3/

REPORT RESUMES

ED 003 077 24
INTERACTING EFFECTS OF VARYING STEP-SIZE AND FEEDBACK IN PROGRAMED INSTRUCTION.
BY- GORDON, JOHN M., JR.
MICHIGAN ST. UNIV., EAST LANSING
REPORT NUMBER CRP-S-066 PUB DATE 65
EDRS PRICE MF-$0.27 HC-$5.84 146P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *EDUCATIONAL RESEARCH, ELEMENTARY SCHOOL STUDENTS, *ACHIEVEMENT, FEEDBACK, SCIENCE COURSES, EVALUATION, SEQUENTIAL LEARNING, INSTRUCTIONAL INTERACTION, EAST LANSING, MICHIGAN

THE FOLLOWING HYPOTHESES WERE TESTED DURING THIS STUDY--(1) PROVIDING KNOWLEDGE OF CORRECT RESPONSE DOES NOT INCREASE THE EFFECTIVENESS OF SMALL STEP-SIZE PROGRAMS FOR STUDENTS AT ALL ACHIEVEMENT LEVELS, (2) PROVIDING KNOWLEDGE OF CORRECT RESPONSE DOES INCREASE THE EFFECTIVENESS OF MODERATE STEP-SIZE PROGRAMS AT ALL ACHIEVEMENT LEVELS, AND (3) PROVIDING MODERATELY DIFFICULT FRAMES WITH KNOWLEDGE OF CORRECT RESPONSE WILL BE MORE EFFECTIVE THAN ANY OTHER COMBINATION OF STEP-SIZE AND FEEDBACK FOR ALL ACHIEVEMENT LEVELS. ABOUT 400 SEVENTH-GRADE SCIENCE STUDENTS WERE ASSIGNED TO TAKE A SELECTED PORTION OF A PUBLISHED PROGRAM COVERING STATIC ELECTRICITY AND VOLTAIC CELLS. THREE VARIATIONS OF THE PROGRAM WERE DEVELOPED UTILIZING DIFFERENT STEP-SIZES. RESULTS WERE EQUIVOCAL. (HB)

ED003077

S-066

INTERACTING EFFECTS OF VARYING STEP-SIZE AND FEEDBACK IN PROGRAMED INSTRUCTION

MICHIGAN STATE UNIVERSITY
1965

JOHN M. GORDON, JR. A THESIS

MICHIGAN STATE UNIVERSITY

[redacted]

ED 003077

INTERACTING EFFECTS OF VARYING STEP-SIZE AND FEEDBACK IN PROGRAMED INSTRUCTION

MICHIGAN STATE UNIVERSITY
1965

JOHN M. GORDON, JR. A THESIS

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

INTERACTING EFFECTS OF VARYING STEP-SIZE AND FEEDBACK IN PROGRAMED INSTRUCTION

Cooperative Research Project No. S-066

John M. Gordon, Director

David R. Krathwohl, Sponsor

Michigan State University
East Lansing, Michigan

1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.

TABLE OF CONTENTS

CHAPTER PAGE

I. PROBLEM.. 1

Problem Perspective.. 1
Rationale and Purpose of the Study.................... 4
  Choosing the Classes of Learners.................. 5
  Choosing the Classes of Objectives................ 5
Theoretical Position and Hypothesis Development....... 7
Related Purpose.. 10
Overview of the Report....................................... 11

II. REVIEW OF THE LITERATURE.................................. 12

Step Size Definition... 12
Effects of Step Size Variation........................ 14
Intelligence and Prior Achievement.................... 18
Effects of Feedback Variation.......................... 20
Cognitive Outcomes... 23
  Long-Term Retention....................................... 23
  Application Transfer...................................... 24
Affective Outcomes... 25
Summary... 27

III. PROCEDURES... 30

Sample... 30
Unique Role of the Four Teachers........................ 31
Content Selection.. 32
Step Size Assessment... 33
  Results of Confidence Rating Assessment............ 34
  Conclusion... 35
Development of Programs of Varying Step Sizes......... 36
Prior Achievement Placement Exam........................ 39
Program Administration....................................... 40
Knowledge and Application Criterion Tests............ 42
Affective Questionnaire...................................... 43
Rationale Underlying the Analyses of the Research.....
  Hypotheses and Questions.................................. 44
Study Design and Rationale................................ 45
Summary... 46

CHAPTER PAGE

IV. RESULTS.. 48

Major Variable Effects Upon Program Characteristics..... 48
  Error Rate.................................. 49
  Time to Complete Program.................... 49
Other Program Characteristics.................. 49
Cognitive Data.................................. 53
  Statistical Rationale........................ 53
  Knowledge-Application Score Correlations, Graphs, and Uniqueness.......... 54
  Multivariate Test............................ 55
  Appropriate Univariate Tests................. 59
  Knowledge Criterion Hypotheses............... 65
  Application Criterion Question............... 68
  Retention Loss Question...................... 69
Affective Data.................................. 71
  Cognitive, Affective, and Cognitive-Affective Correlations.............. 73
Results Summary................................. 77
Discussion Overview............................. 78
Discussion of Problem Section Questions......... 80
Operant Conditioning Position................... 83
Contiguity Position............................. 83
Cognitive Theorists' Position................... 85
  Cognitive Maturity........................... 86
  Entry Repertoire............................. 87
  Individual Accomodation and Pacing........... 88
  Task Difficulty.............................. 90
Overall Conclusion as to Theoretical Rationale.......... 90
New Problems that Evolved....................... 91
Contingent Generalizations...................... 93
Conclusions Concerning the Role of Step Size and Feedback in Programing Strategy.......... 94

V. SUMMARY, CONCLUSIONS, AND IMPLICATIONS FOR NEW RESEARCH.......... 96

Background...................................... 96
Objectives...................................... 97
Procedure....................................... 98
Conclusions..................................... 100
Implications for New Research................... 103

APPENDIX I...................................... 106

APPENDIX II..................................... 119

APPENDIX III.................................... 121

BIBLIOGRAPHY.................................... 133

LIST OF TABLES

TABLE PAGE

TABLE 3.1 EXAMPLE FRAMES FROM THE FOUR STEP SIZE VARIATIONS.... 38

TABLE 4.1 THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE ERROR RATE DATA (N=336)........................... 51

TABLE 4.2 THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE TIME TO COMPLETION DATA (N=336)................. 51

TABLE 4.3 NUMBER OF RESPONSES, DENSITY, PROGRAMING RULES, MEAN ERROR RATES, AND MEAN SECONDS/RESPONSE FOR THE PROGRAM VERSIONS....................................... 52

TABLE 4.4 RELATION OF KNOWLEDGE AND APPLICATION TEST SCORES.... 54

TABLE 4.5 MEANS AND STANDARD DEVIATIONS FOR THE KNOWLEDGE AND APPLICATION SCORES.................................. 56

TABLE 4.6 SUMS, SUMS OF SQUARES AND CROSS PRODUCTS OF KNOWLEDGE (X) AND APPLICATION (Y) SCORES...................... 57

TABLE 4.7 THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE MULTIVARIATE ANALYSIS OF THE COMBINED KNOWLEDGE AND APPLICATION SCORES (N=384)........................... 58

TABLE 4.8 THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE KNOWLEDGE SCORES (N=384)............................ 60

TABLE 4.9 THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE APPLICATION SCORES (N-384)........................... 60

TABLE 4.10 MAIN EFFECT AND INTERACTION MEANS FOR KNOWLEDGE AND APPLICATION TESTS.................................. 61

TABLE 4.11 ANALYSIS OF THE APPLICATION SCORE VARIANCE DUE TO THE ACHIEVEMENT LEVEL MAIN EFFECT THAT REMAINS WHEN KNOWLEDGE SCORES ARE COVARIED OUT................... 62

TABLE 4.12 ANALYSIS OF THE APPLICATION SCORE VARIANCE DUE TO THE STEP SIZE MAIN EFFECT THAT REMAINS WHEN THE KNOWLEDGE SCORES ARE COVARIED OUT.................. 63

TABLE PAGE

TABLE 4.13 ANALYSIS OF THE APPLICATION SCORE VARIANCE DUE TO THE STEP SIZE X FEEDBACK INTERACTION WHEN THE KNOWLEDGE SCORES ARE COVARIED OUT.................. 64

TABLE 4.14 BETWEEN MEAN AND RESIDUAL CORRELATIONS BETWEEN KNOWLEDGE AND APPLICATION SCORES FOR THE SIGNIFICANT SOURCES OF VARIATION................................ 65

TABLE 4.15 KNOWLEDGE SCORE MEANS FOR EACH OVERALL ACHIEVEMENT LEVEL AND TREATMENT GROUP (N=16 PER CELL)........... 66

TABLE 4.16 APPLICATION SCORE MEANS FOR EACH OVERALL ACHIEVEMENT LEVEL AND TREATMENT GROUP (N=16 PER CELL)........... 70

TABLE 4.17 THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE RETENTION LOSS SCORES (N=192)....................... 71

TABLE 4.18 COMPILATION OF THE F-TABLE PROBABILITIES FOR THE SIGNIFICANT MAIN EFFECTS AND INTERACTIONS ON THE AFFECTIVE RATING DATA................................ 72

TABLE 4.19 COMPARISON OF THE INTERESTED - BORED RATING MEANS FOR THE UPPER AND MIDDLE THIRD ACHIEVEMENT LEVEL GROUPS USING THE WITH FEEDBACK VERSIONS (N=14 PER CELL).... 75

TABLE 4.20 COMPARISON OF THE INTERESTED - BORED RATING MEANS FOR THE LOWER THIRD ACHIEVEMENT LEVEL GROUPS USING ALL VERSIONS (N=14 PER CELL)............................ 75

TABLE 4.21 CORRELATIONS BETWEEN ENTRY ACHIEVEMENT AND KNOWLEDGE AND APPLICATION SCORES FOR THE EIGHT PROGRAM GROUPS. 76

TABLE 4.22 COMPILATION OF THE F-TABLE PROBABILITIES FOR THE SIGNIFICANT MAIN EFFECTS AND INTERACTIONS ON MAJOR CRITERION MEASURES................................. 79

CHAPTER I

PROBLEM

The rationale for programed instruction is not universally agreed upon. Consider, for example, the following pairs of contrasting statements. The first of each pair, taken from one of the position papers presented by B. F. Skinner, argues for a particular facet of the technique that bears his name. The second, made by individuals keenly interested in the operations of linear programed instruction, discusses the effects of these facets.

In discussing the need for immediate feedback, Skinner stated that:

> "Immediate feedback encourages a more careful reading of programmed material than is the case in studying a text, where the consequences of attention or inattention are so long deferred that they have little effect on reading skills. The behavior involved in observing or attending to detail — as in inspecting charts and models or listening closely to recorded speech -- is efficiently shaped by the contingencies arranged by the machine. And when an immediate result is in the balance, a student will be more likely to learn how to marshall relevant material, to concentrate on specific features of a presentation, to reject irrelevant materials, to refuse the easy but wrong solution, and to tolerate indecision, all of which are involved in effective thinking" (Skinner, in Lumsdaine and Glaser, 1960, p. 154).

Glaser, after having watched students working through a program, was:

> ". . . impressed that you can leave out, especially in low error rate programs, a lot of the information feedback for many frames, because performing is reinforcing itself and information feedback is often ignored by the student. Sometimes it is necessary to force them to look at it, but that has some instrumentation problems" (Glaser, 1963, p. 188).

Concerning small step size, Skinner wrote:

> "A second requirement of a minimal teaching machine also distinguishes it from earlier versions. In acquiring complex behavior, the student must pass through a carefully designed sequence of steps, often of considerable length. Each step must be so small that it can always be taken, yet in taking it the student moves somewhat closer to fully competent behavior" (Skinner, in Lumsdaine and Glaser, 1960, p. 141).

Lipson, after taking these small steps, remarked:

> "As I proceeded, turning pages and writing down answers, the novelty quickly wore off, and I found myself growing increasingly bored. The steps in the program were so minute that it became less and less necessary to think: the correct answers came with what amounted to dulling certainty" (Lipson, 1962, p. 1).

Finally, Skinner confidently concluded:

> "The learning process is now much better understood. Much of what we know has come from studying the behavior of lower organisms, but the results hold surprisingly well for human subjects. The emphasis in this research has not been on proving or disproving theories but on discovering and controlling the variables of which learning is a function. This practical orientation has paid off, for a surprising degree of control has been achieved. By arranging appropriate 'contingencies of reinforcement', specific forms of behavior can be set up and brought under the control of specific classes of stimuli. The resulting behavior can be maintained in strength for long periods of time" (Skinner, in Lumsdaine and Glaser, 1960, p. 140).

Hilgard surveyed the ensuing research findings and countered:

> "It has turned out that Skinner's confidence, while important in promoting programmed learning, has not been fully sustained. This follows in part because the analogies he uses are just that, and do not necessarily represent identities in the processes involved in Skinner boxes and programmed learning . . . . I believe that the advances made in programmed learning have been based very little upon a strict application of learning theory, regardless of what devotees of the different theories may assert" (Hilgard, 1963, p. 136).

These problems concerning linear programed instruction can be thought of as part of a broader educational question; that of timing, or "when to introduce information into complex, verbal learning to achieve optimal retention, and transfer" (Wittrock and Twelker, 1964, p. 10). Is it, as the linear programed instruction originators stress, most effective and efficient to provide a highly redundant discourse such that there is a high probability that the student will respond correctly, and then follow immediately with the correct response? Or should one provide a less redundant narrative such that the student can only offer a tentative response and then give him the correct answer? Would one strategy be more dependent upon the information within the correct answer than the other?

Moreover, are there different strategies of information introduction for particular classes of learners? It is conceivable that the bright student would flourish with less redundancy and more difficult discourse, whereas the slow student would be retarded.

Learning is also not without its affective counterpart. Are there concomitant affective states which accompany these information strategies that tend to enhance or interfere with the cognitive functioning? Would the less redundant narrative, being more difficult, produce a level of anxiety which would seriously disable certain classes of learners?

The already complex situation is further complicated by the multiple outcomes of instruction, long-term retention and application. Is there a single strategy, using the linear programed instruction framework, that is most effective in attaining both outcomes?

Finally, is there a single position, such as Skinner's, that can provide a rationale for generating solutions to the foregoing questions? Hilgard obviously feels that none of the major positions, that of Skinner, the contiguity theory of Guthrie, or cognitive theory, can account for the "advances made in programmed instruction." If so, is there an amalgam of these positions which would serve as a model for producing operating principles, for programers as producers, and students and teachers as users?

Rationale and Purpose of the Study

Lumsdaine views the problem in these terms and asks "whether a science of instructional programming, dealing with intermediate-level principles, needs to be developed as such - or whether implications for a more general learning and behavior theory can ultimately suffice as a foundation from which a technology of practical programming principles can be derived" (1964, p. 393). He continues to present a rationale for the development of these intermediate-level principles by reasoning that:

> "The need for and probable character of such intermediate level principles in the development of a science of instruction rests in part on the proposition that, in view of the complexity of human learning and the diversity of human learning tasks, we can expect to find relatively few universal generalizations that hold for all classes of instructional objectives, all classes of learners, and all conditions of instruction. Rather what is likely to be most needed is a series of 'contingent' generalizations that take account of the interactions of variables. Experimentally, this position argues for factorial experiments in which two or more variables are studied in combination so that qualifications on a generalization can be determined, and we may validate contingent generalizations of the form: Under condition A, one is obtained, whereas under Condition B, result two is obtained" (p. 394).

The rationale for the present study followed this same line of reasoning and accepted the development of a series of contingent generalizations as its objective. The study proposed a factorial experiment which investigated the interactions between varying amounts of information prior to and after the overt response. Accordingly, these interacting effects were studied with different classes of instructional objectives and different classes of learners.

There might come a point, however, at which some creative individual will combine the principles which this and other research are suggesting into a set of higher-order generalizations - a general learning and behavior theory. As Schramm says, "perhaps this is the curve of progress we should expect - many small advances, resulting, over time, in the accumulation of insights to the size of a critical mass" (1965, p. 9).

Choosing the Classes of Learners

The choice of what classes of learners to sample, as in most educational research, was not really a choice at all. One accepts what is available. Fortunately, four seventh grade teachers in the Lansing schools were interested and willing to try out the experimental materials within their classes. Obtaining seventh graders as a target population was an unexpected surprise, for the achievement range within this grade is quite large. The wide range of achievement made it possible to survey effects of the factorial combinations over three fairly well-defined levels.

Choosing the Classes of Objectives

If we can expect a separate set of contingent generalizations for different educational objectives, then the choice of these objectives and their corresponding criterion performances was crucial. Too few studies have made an attempt to clarify these separate outcomes.

Bloom, et. al., (1956), in their analysis of these various objectives in the cognitive domain, provided a workable framework for separating outcomes. The authors' classification system identified knowledge, comprehension, application, analysis, synthesis, and evaluation as possible categories. Only the first three objectives, knowledge, comprehension, and application, were judged by the teachers in our sample as applicable to the content to be conveyed to their classes. Since the program content was primarily concepts and principles rather than facts, we combined knowledge and comprehension categories into one instrument. A second instrument would be prepared to assess the recognition and application of principles within new situations. Settling upon multiple criterion tests opened the possibility of identifying contingent relations for both outcomes within the one experiment. It also offered to provide evidence regarding the supposed hierarchical relationship of these categories.

Other evidence within the literature pointed to the need for an assessment of long-term retention as a relevant educational objective. Certain of the step size x feedback variations of the program might possibly result in greater retention. It was, therefore, decided to readminister the criterion tests during the following fall using the summer vacation as an interim period.

Krumboltz (1963) introduced more intriguing debate when he contended that program error rate was also a dependent variable. He maintained it was dependent upon the various inter- and intra-frame cue manipulations. What, then, is the relationship between error rate and the other criteria — knowledge and comprehension, application, and retention? The common term for the decoding and encoding of content is acquisition. Is error rate a measure of program acquisition? On the one hand, the answer may be affirmative, in that program frames so closely resemble test items that it could be said that a student making very few errors in working through the program has acquired its content. On the other hand, the answer may be negative, since some would claim that if a student really had acquired the content, he would be able to perform well on a separate knowledge criterion test, even though the test items were almost copies of certain, selected program frames. Still others might say that he really has not acquired the content until he could demonstrate that he had retained the material for some specified length of time. The investigator, consequently, chose to study all forms of acquisition, independently, and in relation to one another.

Theoretical Position and Hypothesis Development

The informational strategy of linear programed instruction calls for a repetitive cycle consisting of: presenting information, leaving one or more words out of context to be filled in by the learner, and immediately offering the missing words. This series is continuously repeated in a sequence which either follows a planned pattern, e.g., ruleg, conversational chaining, or a most logical sequence, that is, logical for the programer. In this seemingly simple informational cycle are all the basic features of the learning process; felt need (to learn), perceived goal (the correct answer), an increase in tension (problem of generating the missing word or words), activity of the provisional try (writing in the perceived words), and feedback (both knowledge of the correct answer and relief of the tension or reward if correct).

The original basic strategy, borrowed from operant conditioning, called for enough prior information to assure the probability of less than one in ten that the learner would make an incorrect response. The learner's response must be written and the correct response must be shown immediately following the overt act. The words selected for the blanks or responses must be crucial to the information content of the program.

Not long after the advent of this strategy in program form, reports of discontent were heard: the frames were boring, writing the response seemed a hindrance, the correct response was being ignored. To explain these occurrences and offer a hypothetical solution, the following set of statements are advanced:

1. If the prior information is so redundant as to make the missing word obvious to the point of certainty, a) there is no increase in tension, hence no challenge, and boredom results; b) there is no need for the information within the feedback; and c) since there is no tension, the feedback loses its reward value as well.
2. If, however, the prior information were to be controlled so that there was a lack of certainty in selecting the response and an accompanying arousal of tension, a) the boredom would disappear; b) the feedback would partially maintain its information value; and c) the choice of a correct response would retain its rewarding function with its reduction of tension.
3. The prior information should not be so precise or so vague that there is no possibility, or only a small probability, that the student will discover the missing words. In such a case, a) too great a tension level is created, b) the information in the feedback is not seen as relevant, and c) its reward value is lost.

Thus, a program whose prior information presentations might evoke both response uncertainty and slight tension, would result in:

1. More efficient content acquisition due to the reduction of prior information redundancy.

2. Assurance that the student complete the full cycle of the learning process which would result in more effective immediate and long-term retention of the program content.
3. A greater probability of problem-solving processes, as yet undefined, being activated which in turn might transfer to tasks where application of content principles is needed for solution.
4. Maintenance of positive affect because of the continuous tension arousal and reduction inherent in this version.

The following research hypotheses and questions developed from these suppositions and related findings of other studies are suggested:

In terms of the comprehension of concepts:

1. Providing knowledge of correct response does not increase the effectiveness of small step size programs for students at all achievement levels.
2. Providing knowledge of correct response does increase the effectiveness of moderate step size programs at all achievement levels.
3. Providing moderately difficult frames with knowledge of correct response will be more effective than any other combination of step size and feedback for all achievement levels.
4. Providing moderately difficult frames with knowledge of correct response will reduce the boredom or "pall" effect among the upper and middle third achievement levels.

The following questions evolved from consideration of other instructional problems that had not been a part of the original rationale.

1. How adequate is the subjective confidence rating the learner gives to his frame response as an indicator of size of step?
2. What effect will programs of varying step size and feedback have upon the learner's ability to apply the concepts and principles within these programs to similar situations?
3. What effect will programs of varying step size and feedback have upon the learner's ability to retain the program concepts over a long interval of time?
4. What relationship exists between prior achievement, program achievement, and the ability to apply the program concepts and principles to similar situations?
5. What effect will programs of varying step size and feedback have upon the level of boredom among the below-average achievers?

Related Purpose

It was initially hoped that the evidence resulting from the study would be applicable to the more common student-teacher instructional interaction. Upon further analysis, there arose a number of distinct differences between the two instructional patterns which make any such inferences quite strained. The following list attempts to relate these major distinctions:

	Student-Teacher Interaction	Linear Programed Instruction
1.	Teacher paced	Student paced
2.	Few students respond	All students respond
3.	More or less sporadic sequence	Empirically tested sequence
4.	Calls for listening comprehension	Calls for reading comprehension

As a result, the study restricts its remarks to situations which relate to the task and methods of linear programed instruction.

Overview of the Report

The discussion within Chapter II centers upon reviewing and relating the findings of other research projects to the theoretical variables and framework just presented. The discussion contains those studies which led to the original hypothesis formulation and also those pertinent studies which have been published since the time of the original proposal.

A further clarification of the methods used to put into operation the main variables, step size and feedback, plus the major procedural decisions are reported in Chapter III. Specific manipulative controlling techniques are explained as well as those in the basic design. A rationale is offered for the statistical analysis that follows in the next chapter.

The specific findings and decisions concerning the theoretical hypotheses and questions are reported in Chapter IV. Summarizing tables are used, wherever possible, to aid in the interpretation of the multiple outcomes. The relation of the results to the theoretical framework is discussed. The report ends in Chapter V with a summary, conclusions, and implications for further research. Appendices and a bibliography follow. The appendices contain graphs of the two-way interactions of the dependent variables, a student by item confidence rating matrix, a sample affective questionnaire, and sample prior achievement, knowledge, and application tests.

CHAPTER II

REVIEW OF THE LITERATURE

The review of the literature will concentrate upon those studies which led to the development of the original theoretical framework. Projects published subsequent to the initiation of the project are included and their relation to the original proposal is explained.

Step Size Definition

The important principle that mastery of a complex subject should be built up by fairly small steps seems no more than common sense. But when Skinner advocated it in conjunction with his other programming rules, the educational world took notice. It was some time before those who accepted these principles realized the inherent complexity in operationally determining the "size" of the steps toward task mastery. Simply calculating the per cent error rate for a program frame and checking this percentage against some pre-determined criterion was little help. It was the actual manipulation of frame content that caused the difficulties. As a result "size of step" began to take on various definitions for example, the number of words per frame, or complexity of length of the response. The meaning associated with error rate, that of degree of difficulty of giving the correct answer, was generally accepted. Yet as Lumsdaine (1959) cogently pointed out this notion of degree of difficulty gave step size a duality of meaning. Difficulty level was a dependent variable, dependent upon the cue manipulation within the frame and the learners on which the program was being validated. The error rate percentage, therefore, was only one indication of the complex interaction between frame content and learner repetoire.

Little progress was made in clarifying this issue. This led Deterline (1963) to state that "no one has yet satisfactorily defined step-size" (p. 6). Smith's solution to the problem was to abolish the term and begin an experimental analysis to "identify the behavioral components of that class of behavior now generally titled 'step-size'" (1963, p. 1). Earlier Shay (1962) had developed a cumulative probability estimate for a given item based on the error rate of all the items in the program. His method, although mathematically and operationally correct, did not help explain the "class of behaviors" that the student performs in working through a sequence of items.

Jacobs (1963) suggested that someone explore the use of a subjective probability estimate of the difficulty of a given frame for a particular learner as a way of assessing this elusive step-size. To determine the reliability of this subjective measure, he surveyed the literature and concluded that "in general, ratings of difficulty, ratings of confidence, and measures of correctness correlated fairly highly in the expected directions" (p. 35-6). Johnson, in a discussion of subjective ratings, said "In all probability, ratings of the difficulty of judgments and the confidence of these judgments may be antonymous designations for the same variable" (1955, p. 168). One objective of the present project was to determine a reliable method for measuring step-size by investigating a method of a assessing the subjective confidence the learner has in his response to a given frame. The confidence measure rather than a subjective probability had been selected on an intuitive basis as being an easier judgment for seventh grade students to make.

Suppes (1964) suggested the use of response latency as a method for defining frame or task performance. He discovered a latency decrease over repeated presentations. Brooks (1964) found median response latency to covary with error rate and concluded,

> "Latency efficiently tells something about all frames -- not just those missed. Error points up excessively difficult frames; latency can indicate as well, those which are too easy. Latency can guide toward greater efficiency or uniformity of task loading, perhaps to arrangements of difficulty which improve student motivation. A quantified measure, the variance of latencies may index the degree of ambiguity. . . . Error and latency data can be complementary aids in developing experimental materials" (p. 4-5).

Effects of Step-Size Variation

Although this research is primarily dedicated to the contingent relationship between step-size and feedback, it is important to review those studies reflecting either of the variables as an individual source of criterion variation. With regard to step-size, the results have been contradictory. Coulson and Silberman (1960) and Campbell (1961) found students working with the small step version scoring significantly better than those using the large step version. Smith and Moore (1962), Briggs (1958), and Shay (1961) found no significant differences. To complicate matters, Goldbeck and Campbell (1962) and Evans, Glaser, and Homme (1960) both obtained results suggesting that some intermediate level may be optimal. All of these studies provided feedback in the form of the correct answer. The possibility of some optimal level above the required ten per cent error rate criterion provided one of the major ties between the literature and the project rationale.

Krumboltz (1963), in a series of clever and comprehensive experiments, attempted a systematic investigation of the effects of various methods of varying step size on error rate and criterion performance. His general thesis was "that [frame] difficulty level is a dependent variable, not an independent variable, and may vary directly or inversely with criterion performance depending on a number of independent variables which influence both difficulty level and criterion performance" (p. 1). This conception sheds a much different light upon the problem of step size variation. In other words, one must be concerned with what it is he is manipulating within the frame, the independent variables, and be aware that step size or difficulty level is an outcome just as are other criterion measures.

These manipulative variables, for want of a better classification, could be split into those that work with a frame -- intra-frame -- and those which manipulate entire frames with respect to the surrounding frames -- inter-frame. One could also think in terms of adding or eliminating information, either intra- or inter-frames. Krumboltz's basic attack seemed to be one of adding information intra-frames.

Adding irrelevant information to already small step frames increased error rate while adding clues and hints decreased error rate as expected. Neither addition, however, affected criterion performance. The basic information, although disguised, remained.

Increasing both error rate and criterion performance was accomplished by asking the students to discriminate between plausible alternatives in a multiple-choice type of program. These results were difficult to compare with findings from the other studies because the task was one of recognition rather than recall. The basic information was, however, supplemented by relevant information within the plausible alternatives. This relevant information probably provided more examples of concepts, repeated more facts, etc., all of which might be expected to increase both kinds of performance.

The present study developed one form similar to Krumboltz's cueing or prompting version in the attempt to construct an easier variation than the basic small step program. In making the program more difficult, both intra- and inter-frame redundancy was reduced by removing information in direct contrast to Krumboltz's addition of irrelevant information, more complex steps, and less familiar terms.

Klaus (1964), while studying the relationship of step-size, error rate, and achievement, defined step-size without reference to learner behavior. Four components; response, cue, context, and enrichment, were used to differentiate frames. Eight indices, four intra-frame and four inter-frame, were developed which employed these four components. A set of procedures, not explained, was used for manipulating frames in the two programs selected for the study. These program frames were analyzed according to a complicated normative system based upon the above indices. Since it was not fully explained, it was impossible to fit his method of information introduction into the overall attack as developed by Krumboltz.

The three step-size versions of both programs had over 700 frames, with 192 subjects participating in the study set up by Klaus. The subjects were divided into three levels by their scores on an intelligence test. Each student was assigned at random to one of the three step-size versions, which resulted in a balanced 3 x 3 (intelligence x step-size) factorial experiment. "No version produced an error rate of over 20%, even among low ability students" (p. 3). Step-size had, however, a significant effect upon the error rate recorded for both programs. Ability level was a significant variable with respect to error rate on one program and not on the other. Klaus gave no indication as to why this difference between programs might have been expected. Both ability level and step-size were significant factors in varying the time to complete the programs.

Klaus failed to find any step-size effects with respect to both proficiency and transfer tests. Ability level did, as one might expect, exert a significant effect upon both criteria. There was, however, no interaction between the two main factors. Step-size variation did not have a differential influence over ability levels. "In summary, the principle findings from the study are that step-size affects error rate, but does not affect achievement and that, when ability level is controlled, error rate and achievement are not significantly related" (p. 5).

It was disappointing that Klaus did not specify his methods of increasing step-size so that one could compare his results within the framework developed by Krumboltz. Furthermore, the differences between error rates for the three versions, although significantly different, did not vary greatly. For example, the mean error rate for the largest step-size version for one of the programs was only 9.4, differing from 7.6 to 12.5 among ability levels. Also, the subjects were considerably above average, with a mean IQ of 117. Greater heterogeneity among the abilities of the learners would undoubtedly have increased error rates, thus making differential main effects and interactions possible.

The present study developed a greater step-size variation over a different age level and more heterogeneous class of learners, seventh graders. Also, the effect of feedback variation was added.

Intelligence and Prior Achievement

In the early literature in linear programmed instruction, Porter (1959), Ferster and Sapon (1958), and Shay (1961) found little relationship between aptitude or intelligence and program criterion achievement. Further studies (Reed and Hayman, 1962; Lambert, Miller, and Wiley, 1962) using longer and more difficult material produced contradictory results in which ability was the major determining factor.

Carroll (1964) found aptitude to be highly related to criterion performance (r = .75), in a study where time to complete the program was held constant. Klaus (1964), as discussed earlier, reported consistent significant differences in criterion performance resulting from initial ability levels. Eigen and Feldhusen (1963) made the most comprehensive study, by assessing the zero and first order partials for IQ, pre-test achievement, reading skills, criterion achievement and transfer. Their general finding was that "reading ability and IQ, while initially correlated with learning, were found to be less essential in accounting for post-test variance than general achievement level prior to work on the programs" (p. 385).

Gotkin and his workers (1964), in a study using seventh grade students quite similar to the present study, maintained that

> "individual differences in pre-test knowledge and cognitive capacity, dictate entirely different programs for different students in the case of most subject matter. . . Even with extensive revisions there is no guarantee that the slow students could attain terminal behavior, and the resulting program would no longer be the one that appealed to bright students" (p. 4).

Gotkin found the culturally disadvantaged students unable to make use of the syntactical prompts, feedback, and redundant frames. He finally had to conclude, "For most of the children, our findings indicate that for seventh graders three years retarded in reading, the ability to learn abstract concepts from printed materials is limited" (p. 4). The solution, he felt, was to develop materials based upon Piaget's concept of developmental stages.

Effects of Feedback Variation

Numerous methods of providing information after the learner's response have been labeled feedback. The term was borrowed from electrical engineers who use it to describe any electrical impulse that serves to regulate the parent system. This lack of a sharply - defined meaning for the term feedback has inevitably led to equivocal findings. Such diverse tasks as target shooting, concept formation, attitude change, and programmed instruction fall within the domain of feedback studies, and therefore, share related results.

Since this study is concerned with programed instruction, an example of highly meaningful verbal learning, the literature review will only touch upon those related studies needed to provide a rationale for our analysis. A distinction first has to be made between the separate, but theoretically dependent, affective and cognitive components of providing feedback. Thorndike's (1913) "law of effect" and Skinner's "reinforcement" (1954) both seemed to center on the reward or affective component of feedback. The effectiveness of a large number of studies, ranging from pigeons playing ping-pong to the learning of paired-associate lists, has been attributed in part to the giving of immediate reinforcement.

It was also recognized that providing knowledge of results was cognitive in nature. The word "knowledge" obviously referred to a capacity to clarify, correct, and confirm one's choice. The cognitive component made its way into the literature through the training methods research dealing with perceptual motor skills.

Ammons (1956), summarizing this literature, found few exceptions to the generally accepted rule that knowledge of results facilitated performance. The tasks, however, were such that the learner's response was primarily a provisional try; a situation where learning the degree and direction of his missing the target was obviously invaluable. The congruity between these tasks and meaningful verbal learning has not been clarified.

Each of these areas, with the main impetus from Skinner, led to the generalization of the feedback principle to programed instruction, even though there was awareness of the task dissimilarity. Smith and Moore (1961) , as well as Hough and Revsin (1963), and Feldhusen and Birt (1962) found no significant differences between programs which did and did not provide feedback. Furthermore, Krumboltz and Weisman (1962) and Lambert (1962) found no differences when they varied schedules of reinforcement or feedback throughout the program. Hough and Revsin (1963) offered the explanation that "when students 'know' that their response is right and thus assumedly reinforce themselves, it would seem reasonable that further confirmation in the form of a reinforcement frame would be redundant" (p.290).

Angell and Lumsdaine (1959) had earlier demonstrated an interaction between the number of prompting trials and of feedback or confirmation/correction panel, as they called it, in a paired associate task. They varied the number of prompting trials, 1, 2, or 3 (which is analogous to manipulating step size), and the method in which feedback was given. The feedback variations were right-wrong, giving a correct and incorrect response, and providing only the correct response. They found a highly significant effect due to prompting trials, while the significant effect due to feedback variations occurred only in the case where no prompting trials were given. Feedback, in any form, became inoperative after two or more prompting trials. In other words, increasing the number of times the subjects experienced the pairs decreased the need for feedback. The extension of these results to redundant cueing within the context of a small step-size program and its corresponding relationship to feedback, provides one of the major generalizations of the study.

Smith and Moore (1962) and Goldbeck and Campbell (1962) also conjectured that there may be a relationship between step size and feedback. Goldbeck, _et_. _al_., in reviewing their studies in Coulson (1962) maintained that:

> "There was too little challenge in the easy items for the overt responding and the formal feedback to become effective. When moderately difficult items were encountered, however, there may have been an increase in motivation and implicit activity associated with the effort for response, with a concomitant increase in the value of the feedback. The increase in the value of the feedback for moderately difficult items appeared to occur whether the overt response was correct or incorrect during the learning trial" (p. 88).

The possibility that there might be an optimal point at which the step size and the accompanying feedback might function for the most effective learning in school subject matter is the basic question of the present study. The increased motivation resulting as a concomitant "side effect" is discussed in the section on affective outcomes.

Further support for the theoretical position was given by Ausubel (1963) who stated:

> ". . . feedback is not generally indispensable for learning,

> but, on both motivation-reinforcement and cognitive grounds, should facilitate the learning process, more so in the case of rote than of meaningful learning. However the research evidence tends to be equivocal, particularly in relation to programed instruction, because of the failure to control other relevant variables. Further compounding the difficulty of interpreting the effect of feedback on meaningful programed learning, is the fact that both low error rate and the possibility of implicit feedback reduce the facilitating potential of explicitly provided feedback" (p. 208).

Finally, Wittrock and Twelker hypothesized and supported the principle that feedback "enhances learning, retention, and transfer, when the information it contains is not greatly redundant," (1964, p. 10) in a problem-solving task involving deciphering ten transpositional cryptograms. They also found an interaction between the amount of direction given within the problem and the amount of feedback. They stated that:

> "It appears that non-redundant information in the form of knowledge of correct response added to a minimally directed situation enhanced learning while redundant knowledge of correct response added to an already prompted situation, did little to learning, retention, and transfer" (p. 17).

This study was reported after the initial proposal and represents an intermediate link from Angell and Lumdaine's paired-associate task finding to the proposition that this interaction is also apparent within the learning of school subject matter.

Other Cognitive Outcomes

Long Term Retention A retention test of some form has been part of many studies in programed instruction. Strong (1963) summarized these by stating, "Most of the research suggests that programed instruction maintains its superiority over time but it is not superior in terms of percentage retention" (1963, p. 226). Krumboltz and Weisman (1963) demonstrated some contrary evidence as they found greater retention occuring when students were forced to give overt responses over a two-week interval.

Application-Transfer Gagne and Dick (1962) included a transfer test in a study attempting to "measure and define the nature of 'what was learned in a teaching-machine program on solving simple algebraic equations of the first order'" (1962, p. 10). They were attempting to demonstrate Gagne's contention that the success at a higher task level was primarily dependent upon the acquisition of specific and relevant lower order tasks. The students performed very poorly on the transfer test, scoring a mean of 2.07 out of a possible total of 50. The restricted amount of variance about the mean of 2.07 made the reported correlations between transfer scores and the other variables questionable. Wendt and Rust (1963) demonstrated the use of pictorial frames as beneficial for transfer to the real life situation.

Klaus (1964), studying the effects of varying step size over ability levels, reported no significant differences between transfer scores due to step-size. His step-size variations were not very large, however, with only one version having an overall mean error rate over 10 per cent. He did find significant differences resulting from the initial ability placement. Unfortunately, correlations between ability, post-test, and transfer scores were not reported.

Eigen and Feldhusen, in a series of studies reported in DeCecco (1964), found evidence for Gagne's hypothesis that "achievement at one stage is a principal determiner of success in learning at a next higher stage of learning new material" (p. 384). In studying the interrelationships among reading ability, IQ, post-test acquisition, and transfer, they stated that "it is inferred from the consistently high correlations of acquisition and transfer scores that successful 'performance' in a new class of tasks for which the learner must make adaptations is primarily dependent upon mastery or achievement of a subordinate set of learning tasks as measured by the acquisition tests" (p. 384).

Affective Outcomes

> "As I proceeded, turning pages and writing down answers, the novelty quickly wore off, and I found myself growing increasingly bored. The steps in the program were so minute that it became less and less necessary to think; the correct answers came with what amounted to dulling certainty" (Lipson, 1962, p. 2).

Affective reactions to programed instruction have been reported which vary from the ridiculous to the sublime (Naumann, 1962, Roth, 1963, Eigen, 1963). Among the former have been the recurring statements which refer to a form of boredom or monotony which is said to result from highly redundant information rate. This deleterious side effect has been given the unenviable label of the "pall" effect.

Goldbeck suggested in an earlier quote that "when moderately difficult items were encountered, however, there would be an increase in motivation." McClelland, Atkinson, et. al. (1953), in their classic study of achievement motivation, posited this theoretical explanation for the occurrence of positive and negative affect which reads as though it were an attempt to define the situation stated by Goldbeck.

> "If the expectation is of low probability [large step-size], then confirmation should produce negative affect as in 'fear of the strange.' If they are of moderate probability [moderate step-size] precise confirmation should produce

> pleasure as in reading a detective story or playing solataire. If the expectations are of high probability [small step-size], precise confirmation produces boredom or indifference" (p. 87).

This explanation provides the theoretical rationale for the hypothesis of reduced boredom or 'pall' effect from the moderate step-size versions of the programs for the above average and average groups. What will transpire with the below average group, which had little confirmation throughout their schooling, remains a question.

In a related study, not using programed materials, Chansky (1964) found that "attitudes students express about learning seem to be related to the acquisition phase of learning but not to the retention phase" (p. 99). Students, given four methods of instruction for learning a given task, rated those giving continuous feedback as most interesting and intermittent grading more worry-provoking. The less provoking methods did not, however, result in better retention. Intermittent grading was most effective and efficient in terms of retention. The comparison to step-size and feedback variation as producing affective change and its effect upon cognitive outcomes, hopefully, is obvious.

Eigen and Feldhusen, as reported in DeCecco (1964), found "that students' attitudes toward the program are not generally correlated with their success in learning or transferring from the program" (p. 385). No interpretations were given as to why the correlations were not discovered.

Summary

The definition and measurement of step-size has been a constant source of trouble for those advocating linear programed instruction. The assessment of either frame or program error rate offered some information as to the relative difficulty, the most common meaning, but little understanding of the complex interaction between student and frame cues which produced the response. Lumsdaine, as early as 1959, promoted a concentrated attack upon the cue manipulation intra- and inter- frames to offer some evidence as to the best methods for creating small step size. Jacobs (1963) called for a better measurement, in the sense of giving more information than the dichotomous right-wrong. He suggested the use of response latency, difficulty ratings, and a confidence estimate. Suppes (1964), and Brooks (1964) experimented with response latency and found that it complemented the dichotomous error rate. The present study investigated the use of a confidence estimate.

Contradictory results were found regarding those studies dealing with step-size effects. This is not surprising when one considers the many ways in which step-size can be defined and varied. Krumboltz (1963) considered the error rate meaning of step size to be a dependent variable and conducted a number of experiments carefully manipulating frame characteristics. In most cases, he added different forms of information to the already low error rate program. This additional information generally influenced error rate but not criterion performance. In contrast the present study both added and removed intra- and inter-frame information.

Klaus (1964) talked of a complex operation to manipulate step size without reference to the learner. Unfortunately he did not specify exactly how the information was removed. He found, however, that his manipulations resulted in only small error rate differences. These differences were significant but had no effect upon criterion performance.

Early studies, dealing with a program having a small number of frames, revealed no interaction between program variables and student ability. Later, Carroll (1964) controlled for time to complete the program and found a large criterion-ability correlation. Eigen and Feldhusen (1964), however, continued the attack and found both IQ and reading ability washing out when prior achievement level was correlated with criterion performance.

A study by Gotkin (1964), which discussed the unique problems of programing for culturally deprived seventh grade students, was discussed because of its similarity to the present study sample. His difficulties and conclusions mirrored those of the present study.

The concept of feedback has experienced much the same semantic confusion as has step size, mainly due to its dual meaning. Feedback inherently has both a cognitive, informative value and affective reward value. Many studies, purporting general statements concerning feedback, however, have not recognized the duality.

Studies emphasizing the cognitive aspect have centered upon difficult tasks, difficult enough that the student's first responses were termed "provisional tries." This situation is quite unlike the relative certainty one has in responding to small step program frames. Researchers have used this explanation when they began to find that withholding the correct response had no adverse effect upon criterion performance.

Angell and Lumsdaine (1959) found a prompting trial confirmation interaction in an experiment using a paired-associate task. The question as to whether a concomitant effect could be demonstrated within complex meaningful verbal learning was dropped by Goldbeck (1962).

Programed instruction's effectiveness in bringing about change in long-term as well as immediate retention of knowledge has yet to be firmly demonstrated. Only Krumboltz and Weisman (1963) have found some indication of greater retention with overt responding, within a two-week interval. The present study investigates the amount of retention over the typical summer vacation.

The ways to facilitate transfer or applicationof concepts and principles to new situations are also little understood. Eigen and Feldhusen (1964) did find evidence for Gagne's contention that prior achievement and program acquisition are correlated with transfer performance.

A boredom called the "pall" effect, occurring from the small-step monotony, has received much attention as an outcome of small step-size programs. Chansky (1964), however, found no deleterious criterion performance effects due to worry-provoking methods. Eigen and Feldhusen (1964) found similar results from a long-term study in programed instruction, that is, no correlation between attitude and performance. Goldbeck (1962) suggested that more difficult items might have a positive influence upon motivation. His hunch is backed up by a number of psychological studies, most vividly by McClelland and Atkinson (1953).

CHAPTER III

PROCEDURES

Overview

The chapter begins with a description of the sample and the research setting, followed by an elaboration of the role of the four science teachers. Next, the selection of program content and step-size estimation using the confidence ratings are discussed. The results and conclusions surrounding the use of the confidence ratings are then presented, which lead to the final development of the step-size variations. Criterion instrument development techniques and reliability estimates are given. Finally, the design and statistical analysis are explained.

Sample

Permission was granted and assistance given by the Lansing, Michigan school officials to carry out the study at West Junior High School of that city. Four seventh grade science teachers, in particular, expressed interest in the use of programed materials. They were especially concerned about meeting the needs demanded by the large number of low-achieving students.

A total of 16 classes, all but one belonging to these four teachers, were used in study, bringing the total to over 500 students. Although they constituted neither the total seventh grade population of 570 nor a bona fide "probability" sample, it was the teachers' judgment that those students taking part in the study were representative of the entire seventh grade.

West is the most central of the junior high schools of Lansing. It draws its students from all cultural, racial, and economic groups. A large segment of the children comes from a residential, upper-middle, and upper-class district populated mostly by professional people employed in managerial positions at Oldsmobile and White Motor Car Corporations. A similar segment of the children resides in low- and lower-middle class district. If anything might be considered abnormal about this particular school population, it is that it is lacking the common majority in the middle representative of the middle peak on the normal curve. By dividing the overall group in thirds according to prior science achievement, it was possible to study each segment independently as well as the total group.

A reasonable estimate of the variability of these students on certain skills, such as reading, would be from second to eleventh grade levels. This wide variability was both a help and hindrance. It was a help in the sense that it provided information of the program effects over all levels, and a hindrance in that developing instructional materials that were adequate for all students was nearly impossible.

Unique Role of the Four Teachers

The four interested teachers were significantly involved in every major decision, except those concerning basic design and analysis. They received an honorarium for this work. As a group they made the major curricular decision, that of what the program content should be. Three were given portions of the basic program to revise while a fourth developed criterion test items. Their combined judgment concerning the basic wording of the textual material, the student readiness for certain concepts, and whether or not the students had acquired prerequisite and program concepts was accepted. Most important were their assessments and aid in developing the directions for the program and criterion instruments. Each set of directions was tested before program administration to assure its comprehensability, but it was the teachers' pre-judging that greatly facilitated the task.

Content Selection

What was thought to be a simple task, that of choosing the program content, was, in fact, a most difficult and important decision. Although every teacher was using the same textbook, each of their classes was at a different point in the curricular sequence. The final selection, a unit about static electricity and voltaic cells, was normally considered a topic to be covered in the following year. Fortunately, the prerequisite concepts, those under the heading of magnetism, were a part of the sixth and seventh grade content. In addition, static electricity was not covered by the seventh grade text which eliminated much of the problem surrounding individual home study. Yet there remained the problem of readiness, that is, were these students conceptually ready for the abstract nature of the chosen content.

Only one nationally published program could be located which covered static electricity and voltaic cells. Two hundred and twelve frames within General Science, Sound, Sight, Electricity, and Communications, Vol. II, by Schaefer, Jeffries, Phillips, Harakas, and Glaser, distributed by Teaching Machines Corporation, dealt with these topics. The number of frames chosen was considered large enough to: 1) eliminate the short program complaints and 2) make it possible for the pall effect to occur.

The immediate reaction of the teachers was that the program wording was too difficult for the low achiever. That decision prompted a major revision centering on bolstering weak portions and removing difficult examples. Two complete revisions, each using representative students as a source of validation, were carried out to reach a satisfactory version. The first revision was tested upon a heterogeneous class of thirty students while the second was validated upon eight low achieving students. These students were excluded from the group who participated in the experimental phase of the project.

Step-Size Assessment

The adequacy of the confidence es[illegible]te as a more informative assessment of step-size was tested during the revision phase. The first task was to develop a set of directions and a rating system which could be easily understood and performed even by the poorest reader. After a number of trials, mainly with low achieving students, the following graphic and verbal directions were settled upon.

> After writing your answer to each question on the separate sheet of paper, write down the number which describes how confident you are of your answer.

0	1	2	3	4	5	6	7	8	9	10
no idea		wild guess		good guess			pretty sure			certain

The first revision group, the heterogenous class of thirty, was asked to work through the program and to respond, both to the frame, and with their confidence in that response. It was hoped that certain frames would receive clusters of confidence estimates around the middle or five rating. These frames would then be analyzed as to their unique cueing qualities and remain in the program. The cueing properties of the remaining frames were to be adjusted in line with these unique cueing methods. Frames receiving primarily high ratings would be made more difficult while frames receiving low ratings would be made easier.

A student by frame confidence rating matrix was generated to easily assess overall ratings. (See Appendix II) The students were ranked as to their science achievement by their instructor, while the frames were in the program order. Each rating was recorded and circled if the response was incorrect.

Results of Confidence Rating Assessment Although it is common in research studies to record results in a separate chapter, decisions concerning these ratings influenced the development of the step-size variations. To explain the methods finally used in manipulating step-size it is necessary to discuss this phase's results here.

None of the expected clusters of ratings between three and seven appeared among any of the frames. The ratings tended to be high or low. The students were either confident they knew the correct answer, or were confident they did not know the answer. Less than ten per cent of the ratings fell into the "good guess" range. Upon further inquiry, it was learned that one bright boy rated a frame at five because he felt it was "ambiguous".

Low achieving students tended to mark eight's and nine's in contrast to the better students, who rated mostly ten's. It seemed that low students were not able to admit to "certain" even on many of the easiest items. This might have been expected, as theirs is a history of failure on tests which the programs greatly resemble.

Conclusion

This type of rating offered only limited information beyond the accumulation of errors and corresponding error rate. Therefore, the ratings proved of little use in determining the frame factors which might account for moderate difficulty levels. Nevertheless, the part confidence or certainty plays in the meaningful learning is still an interesting problem that should be explored in further research.

While observing individual students, it was not uncommon to find one pondering some time on a frame, suddenly write down the answer, and rate his confidence in that answer as ten, "certain". This activity suggested that some method for determining response latency, as advocated by Brooks, Suppes, and Jacobs, seemed to be appropriate as a more differentiating measurement for step size.

Since the school year was drawing to an end, it was agreed that to develop the latency method and run more trials would take the study into the summer. A more arbitrary attack on step-size had to be taken.

Development of Programs of Varying Step Size

An arbitrary but systematic way of manipulating step size was needed. First, the basic program revision was continued until the requirement of small step size, below ten per cent error rate, was met with the low achievers. More frames were added until a total of 278 with 330 separate responses were written. This final version was administered to eight low achievers and their combined error rate was twelve per cent. Twelve per cent was considered close to the ten per cent criteria so one more effort was made. It was imperative that the original or basic program be within the accepted limits. Following Krumboltz's lead, it was decided to add the first one or two letters of the correct response within the frame blank. Different sections of frames were tested in short trial runs and found to be satisfactory for reducing errors.

Two more difficult versions were needed to sample the step size continuum. More than two would have taxed the design by increasing the number of cells and thereby reducing the number of students within each cell. It was imperative, however, to retain the basic information so that the programs were not teaching more or less facts, concepts, and principles. Krumboltz's attack was largely one of increasing the basic program by adding both irrelevant and redundant information. Therefore the opposite strategy, removal of words, was accepted. Any deleted words and frames had to be considered redundant or repetitive. What might be redundant and repetitive for one learner, however, might not be for another learner. Any removal of the words and frames considered redundant had to be somewhat arbitrary.

The decision was made to remove both inter- and intra-frame redundancy thus insuring that step size would be measurably increased. Inter-frame redundancy was reduced by eliminating both redundant frames within a group of frames that introduce new information and review frames that are far removed from the initial information introduction. Intra-frame redundancy was removed by eliminating selected words within the frame which serve to cue the response. Both techniques were designed to increase the difficulty of generating the correct response.

Each frame of the basic program was then classified by the experimenter as to whether it contained new information, was redundant, or was review. The strategy for developing the "moderately" difficult program, hypothesized to be both most effective and best able to demonstrate operative feedback, was to have one associated redundant frame, and two related review frames removed, plus one contextually redundant word deleted from each remaining frame. The final, "most difficult" variation had two associated redundant frames and all review frames erased, plus two contextually redundant words from each remaining frame. (See Table 3.1)

The underlying rationale was to reduce: 1) inter-frame redundancy to increase the pace thus negating the possibility of the pall effect due to repetition, and 2) intra-frame redundancy

TABLE 3.1

EXAMPLE FRAMES FROM THE FOUR STEP SIZE VARIATIONS

Variation I. With Feedback

20. The zinc plate reacts with the solution and causes zinc atoms to leave the plate and go into the s__________.

solution

21. After these atoms in the solution leave electrons on the remaining plate, they have more protons than electrons. The atoms then have a p__________ electrical charge.

positive

Variation II. Without Feedback

20. The zinc plate reacts with the solution and causes zinc atoms to leave the plate and go into the __________.

21. After these atoms in the solution leave electrons on the remaining plate, they have more protons than electrons. The atoms then have a __________ electrical charge.

Variation III. With Feedback

14. The zinc plate reacts with the __________ and causes zinc atoms to leave the plate and go into the _______*.

solution

15. After these atoms in the solution leave electrons on the remaining plate, they have more protons than __________. The atoms then have a __________* electrical charge.

positive

Variation IV. Without Feedback

9. The zinc plate reacts with the __________ and causes __________ atoms to leave the plate and go into the __________*.

10. After these atoms in the solution leave electrons on the remaining plate, they have more __________ than __________. The atoms then have a __________* electrical charge.

* denotes the desired response

to increase the individual task difficulty, thus adding to the possible tension arousal and subsequent relief. The moderate level was to maximize both difficulty and pace so as to best suit the majority of learners. The final variation was expected to be beyond the difficulty level and pace of most of the students, thus causing a frustration from little or no tension relief as well as cognitive confusion.

The number of ways feedback could be varied was also seriously limited by the number of cells in the factorial experiment. To insure adequate within-cell estimation, only two variations would be allowed. It was natural to choose the two extremes, with and without knowledge of correct response. Therefore each of the four step size variations was mimeographed twice, once with, and one without, the correct response. These combinations brought the the total number of different versions to eight.

Prior Achievement Placement Exam

The theoretical hypotheses were generated in the expectation that students judged to be achieving in the upper, middle, and lower third of their class would be differentially effected by the varying step size and feedback program combinations. It was therefore necessary to administer a published or teacher-constructed test to determine in which third to place each student. The evidence of Eigen and Feldhusen (1964), who found prior achievement to be correlated with criterion performance even when IQ and reading ability were partialled out, supported the choice of prior achievement rather than other ability measures.

There still remained the choice of whether to use a pre-test consisting of items testing the program content or one which assessed the overall progress to that point. The inherent difficulty arising from the use of a pre-test, of differential pre-test x treatment interaction over levels, determined the choice of the overall progress test. A search through the published science achievement tests did not produce a test which corresponded with the teacher's curricular approach. There was concern over the lower achievers being able to answer some of the items. It was decided to sample equally from the best items of previous exams constructed by the four teachers plus extracting other items from the teacher's manuals of relevant texts. Forty-one items were chosen by the teachers which would not favor any class or level of achiever.

The test was administered approximately two weeks prior to the program administration. As expected, the test scores developed a wide variation, forming a platykurtic distribution which denoted the lack of the usually prominent middle group. It was therefore easy to rank the scores and divide the students into upper, middle, and lower thirds representing the desired levels of prior achievement.

Program Administration

The programs and corresponding criterion instruments were administered by the classroom teachers during the daily, 50-minute periods, during the last two weeks in May, 1964. Time to complete the instruments ranged from two to eight class periods before the slow students were finished. Eight students did not finish because of absence and inability to read. Students finishing early were given independent remedial or enriching study.

The teachers were given a standard set of directions to give to each class. These instructions included separate statements for those receiving each feedback form and special statements for those receiving the two more difficult forms. The teachers remarked that each class needed extra emphasis and repetition, especially the "modified" classes consisting of low achievers. The fact that everyone had different booklets was only distracting at first. The teachers also gave individual assistance when needed.

The teachers told their students that they were experimenting with a new kind of textbook to eliminate questions dealing with the nature of the task. It was not uncommon for these teachers to conduct experiments of this kind in their science classes. They also instructed the students to be prepared to answer test questions that would not, however, be graded, but would be part of the material included in the final exam. It was hoped that these conditions would produce maximum motivation without inducing outside study and other detrimental influences.

There were no signs of student discontent with the task which led two teachers, independently, to remark that "it was the quietest the room had been all year." The teachers also kept records of the amount of time needed by each student to work through the program. These time estimates became an integral part of the data analysis.

As soon as the student finished the program, he rated himself on the continua representing the extremes of the selected adjective pairs. Upon completing the questionnaire, he was given either or both of the criterion tests depending upon the time remaining.

In October of the next school year, the school officials allowed one period, fifty minutes, for retesting the students, now eighth graders. Both criterion tests were administered but many of the students failed to complete the application portion. Only the knowledge scores were used in the data analysis.

Knowledge and Comprehension and Application Criterion Tests

The cognitive criterion instruments were patterned after the knowledge, comprehension, and application categories as set forth in Handbook I of the Taxonomy of Educational Objectives. Thirty different facts and concepts were culled from the program. Each member of the research team developed a set of representative items calling for either retrieval of the fact or comprehension of the concept. Items, judged to be most discriminating by the consensus of the teachers, were each given a weight of one to form a thirty item completion exam. (See Appendix III) An estimate of the reliability of the test, given by the Kuder Richardson Formula 21, was .87.

The application items (See Appendix III) were developed from experiments and practical problem situations whose solutions required a knowledge of the conceptual meaning and skill in applying the principles within the program. The student was asked to name the principle involved and the correct application of that principle. Eighteen items were constructed, each worth two points, one for correct elaboration of the principle and one point for the correct application. It was felt that the large achievement levels effect found in the analysis of the application score variance provided good evidence of the test's internal consistency.

Affective Questionnaire The affective questionnaire consisted of six adjective pairs similar to those in the Semantic Differential (see Appendix III). The numbers, one to nine, were added because the low achievers had difficulty in comprehending the directions. Equal interval assumptions were made to simplify the computations.

The adjective pairs were chosen in an attempt to cover the possible relevant facets of the affective response. The investigator searched through the literature on attitudinal response to programs and through Roget's Thesaurus before selecting the following pairs: difficult-easy, alert-careless, rewarded-punished, progress-no progress, successful-frustrated, and interested-bored. The final pair, interested-bored, was used to test the hypotheses concerning the reduction of pall effect.

Since the design called for each student to study from only one of the eight forms, he obviously was not able to make any comparative judgments among programs. The affective data was single stimulus data and therefore much weaker than if the students had been able to contrast each of the forms.

The Rationale Underlying the Analyses of the Research Hypotheses and Questions

Although the specific research hypotheses dealt with individual criterion measures independently, it was possible because of the multiple scores on one student to consider the use of multivariate analysis of variance. The multivariate technique enables one to make decisions regarding a combination of dependent measures as if they were a single measure. The completion of these tests offered an overall view of the variance dispersion. Since it was only feasible to compute the multivariate ratios for a problem having two dependent measures on a desk calculator, only the prime variables, knowledge and application, were analyzed in this fashion.

MultiANOVA considers the dependency of the criterion test scores in making an overall test of the differences obtained from the main effects and interactions. One is not allowed, for example, to make independent probability judgments upon the main effects of feedback on application scores or knowledge scores without regard for their possible covariations. The assumptions of this omnibus test are multivariate normality and homogeneity of variances and covariances. Oddly enough both of the tests for the assumptions are more complex than the between means test. Fortunately wide deviations within the data which would cause rejection of the assumptions, were easy to spot when calculating the variances and covariances.

The resultant variance-covariance ratios, W's, are distributed as Chi-square for a certain number of degrees of freedom. One obtains eight W's corresponding to the three main effects, three two-way interactions, one three-way interaction, and the overall between means effects.

Hypothesis #1, stated in the negative, needed a reversal, that is, without feedback versions being more effective, to be logically substantiated. A priori hypotheses #2, #3, and #4 were to be tested by specific series of t-tests within levels plus an over levels test.

The research questions called for an after-the-fact analysis of the variance components of the dependent measure associated with each question. The major technique was a three-way factorial breakdown with step size, feedback, and achievement level as the main effects along with their interactions. Some affective-cognitive correlations were also under study to guage the possible relationships.

Study Design and Rationale

The scores on the placement test were ranked and divided into three groups to represent the three prior achievement levels. Starting with the top eight scores, each student was randomly assigned through use of the random numbers table to one of the eight program variations. Each succeeding group of eight ranked papers was similarly assigned until all the students were placed. The result was a two-way (4 x 2) factorial experiment randomized within blocks design, having sixty blocks each containing eight students. The two factors represented the four step size variations and two types of feedback while the blocks were derived from the prior achievement scores.

After the samples were selected it was decided to collapse the "blocks" into three levels for the statistical analysis in order to study the overall levels effect and the various main effect by level interactions. The optimal statistical treatment would have been to retain the sixty blocks which would remove the maximum amount of "true" variance. The alternative procedure, however, was thought to be defensible on the basis that it would result in a more conservative estimate of error variance than would have been produced from a sampling plan based upon a randomization within levels design.

Assuring that each of the eight program combinations be representated in each block, within the three levels, would have the effect of increasing the within group or error variance over that which would have resulted from the simpler randomization within levels.

Summary

The manner in which the student sample and cooperating school officials were chosen was discussed initially. Next the role and delegated responsibilities of the four teachers in selecting the content, advising on wording, etc. was reviewed. Procedures for assessing and analyzing the subjective confidence ratings followed, concluding with the statement that the ratings offered little information beyond the dichotomous error rate.

Most of the ratings were either high or low while desired or middle ratings were said to be associated with ambiguous items. More work needs to be done to explain the function of certainty and confidence in problem solving.

Different kinds of information pointed to the response latency measure as the most promising attack upon step size definition.

The resulting strategies for information reduction to increase step size were elaborated upon and examples presented. Criterion instrument development and program administration details were explained. Lastly the rationale for the design and hypothesis testing was offered.

CHAPTER IV

RESULTS

Overview

The efficiency of the factorial experiment with multiple outcomes is clearly demonstrated by the length and complexity of the following chapter and corresponding appendix. The major form of statistical analysis is the three-way factorial source of variation breakdown of the three main factors and their interactions. The significant effects are starred in the tables, mentioned in the text, and summarized in a final table at the end of the chapter. The overall effects might be best understood by glancing at the summary table first.

The effects of the major variables upon the program measures are presented first to maintain some continuity with the actual collection and analysis. The multivariate and acceptable univariate analyses of the cognitive variables, as a group, follow. The theoretical hypotheses and questions are taken up in their original order. The variable effects upon the affective ratings, as a group, are printed next, with the appropriate hypotheses and questions being answered.

An exploratory look at the more significant relationships between and among the cognitive and affective data precedes the summary of the findings.

Error Rate

The error rates have to be interpreted in light of the failure of some students to heed the directions which called for making a provisional try before looking at the correct answer. The major effect of this response peeking was to make the lower third estimates too low. (See Table 4.3). The overall effect upon the source of variance breakdown as shown in Table 4.1, was difficult to discern. The overall strength of the effect of the main variables in influencing error rate showed that every main effect and interaction produced a significant variation. Interaction plots are found in Appendix I, Graph #3. If the program frames were to be considered the first "learning trial", then each main variable, entry achievement level, step size, and feedback, and all of their interactions have a pronounced effect on the outcome.

Time to Complete the Program

Every main effect and interaction, except for the three-way, had a significant effect upon the time to complete the program (Table 4.2). Interaction plots are found in Appendix I, Graph #4. The program manipulations all resulted in varying the students' pace through the program information.

Other Program Characteristics

Other program characteristics such as density, mean seconds per response, mean total minutes to complete the program, as well as error rate, are depicted in Table 4.3. Program density is the ratio formed by the number of different responses over the total number of responses. Density measures give a quick assessment of the amount of redundancy, and, therefore, difficulty, of the program. They are unique in that they are independent of the performance of the students. Variations I and II remained at the same density ratio, .21, because nc frames were removed. Reducing inter-frame redundancy on Variations III and IV decreased the density to .33 and .43, respectively. A total of 217 (330-113) frames were removed, 21 (70-49) of them having different responses than the original group. These 21 items were judged to contain the same information as previous frames but called for a secondary response, usually an adjective.

A gross measure of response latency was also taken, that of the mean number of seconds taken per response. This estimate was calculated from the mean minutes to complete the program, also shown. The mean seconds per response estimates demonstrate an interesting phenomena. The with-feedback versions of Variations III and IV were almost equal, 39 to 40, while the without feedback versions deffered greatly, 43 to 59. The presence of feedback considerably decreased the number of seconds before responding on Variation IV. It is conjectured that the presence of feedback made the students "give up" faster as the anxiety of wondering about the response increased. The similarity between Variation III and IV, 39 and 40 mean seconds, seems to point to a kind of tolerance limit. The without feedback group of Variation IV took, on the average, almost 20 seconds more per response. They were willing to work longer to find the solution. This extra time taken could have been the major factor in the increase of application scores to be exhibited in the next section.

TABLE 4.1

THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE ERROR RATE DATA (N=336)

Scource of Variation	Sum of Squares	d.f.	Mean Square	F
Achievement Levels	16,515	2	8,257	120.19***
Step Size	55,475	3	18,491	269.17***
Feedback	46,008	1	46,008	669.69***
Feedback x Step Size	20,194	3	6,731	97.98***
Achievement Levels x Step Size	1,710	6	2,618	38.11***
Achievement Levels x Feedback	15,289	2	7,644	111.27***
Achievement Levels x Feedback x Step Size	3,857	6	642	9.35**
Within Cells	21,504	313	68	

*** p <.001 ** p <.01

TABLE 4.2.

THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE TIME TO COMPLETION DATA (N=336)

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Achievement Levels	31,576	2	15,788	17.2***
Step Size	96,509	3	32,169	35.1***
Feedback	42,525	1	42,525	46.5***
Feedback x Step Size	13,979	3	4,494	4.9**
Achievement Levels x Step Size	20,632	6	3,455	3.7**
Achievement Levels x Feedback	7,380	2	3,690	4.0*
Achievement Levels x Feedback x Step Size	5,378	6	896	
Within Cells	285,342	312	914	

*** p <.001 **p <.01 *p <.05

Table 4.3

NUMBER OF RESPONSES, DENSITY, PROGRAMING RULES, MEAN ERROR RATES, AND MEAN SECONDS/RESPONSE FOR THE PROGRAM VERSIONS

Step Size Variations	Number of Responses	Density*	Programing Rules	Mean Error Rates			Mean Secs/Response	
					w	w/o		
I.	330	$\frac{70}{330} = .21$	Prompt All	UT MT LT	2 4 8	8 12 23	w w/o	21 23
II.	330	$\frac{70}{330} = .21$	Master Program	UT MT LT	6 11 7	14 23 64	w w/o	23 25
III.	179	$\frac{59}{179} = .33$	Delete - 1. One Redundant Frame 2. Two Review Frames 3. One Key Word/ Frame	UT MT LT	12 17 16	42 44 72	w w/o	39 43
IV.	113	$\frac{49}{113} = .43$	Delete - 1. Two Redundant Frames 2. All Review Frames 3. Two Key Words/ Frame	UT MT LT	24 25 29	40 60 82	w w/o	40 54

Key:
UT = Upper Third
MT = Middle Third
LT = Lower Third
w = with feedback
w/o = without feedback

*Total Number of Different Responses/Total Number of Responses

There was a marked similarity in the between-variation increments of density, response latency, and error rate measures. The inability to reflect intra-frame redundancy makes the density ratio a more gross measure, but the fact that it can be calculated before the program is administered recommends it highly. The expected response-peeking done by the with-feedback groups seriously hinders interpretation of the error rate data. The without feedback versions error rates, where there was no chance of peeking, are a more reliable measure but do not reflect the "normal situation," that is, normal programs contain feedback. The gross response latency measure probably most clearly represents the student reaction to the difficulty of the programs. This conclusion agrees with the earlier work with the confidence rating.

Cognitive Data

Statistical Rationale With dependent measures, the obvious statistical test to run would have been one grand multivariate test covering the overall variable effects on the multiple criterion measures: error rate, completion time, the seven affective ratings, the knowledge and comprehension scores, the application scores, and the retention loss scores. This omnibus test takes account of the natural covariance between criterion measures since they are measures on the same subjects. It was impossible, at the time, to carry out the complete, overall test because of the lack of a suitable computer program. However, the two major criterion measures, the knowledge and application scores, were put to the multivariate test. Kendall (1961) provided the model for the two-measure case which could be done in a reasonable amount of time using an electric calculator.

The decisions reached within the overall multivariate tests are similar to the more common univariate except that they consider all the criterion measures as one combination. For example, we wish to make major programing decisions on the basis of the overall effects upon both objectives, knowledge and application, as though it were a single score. In the present study the overall test was used to demonstrate its relation to the individual tests which represented the decisions needed to answer the research hypotheses. The individual hypothesis tests could have been made independently of the overall tests because of their a priori status. The general logic from the multivariate to these individual tests happened to be consistent and was reported in this manner. Of interest in multivariate tests is the relationship between dependent variables among the treatment groups. This analysis follows.

Knowledge-Application Score Correlations, Graphs, and Uniqueness

The overall Pearson product moment correlation coefficient between knowledge and application scores was .76. The coefficient remained quite stable over program variations. (See Table 4.4)

TABLE 4.4

RELATION OF KNOWLEDGE AND
APPLICATION TEST SCORES

Step Size Variations								
Feedback Versions	w	w/o	w	w/o	w	w/o	w	w/o
Correlations	.76	.73	.65	.70	.66	.74	.81	.80
n's	42	40	45	39	44	38	46	44

The resulting scattergrams were unique in that their points concentrated in the lower right diagonal half. (See Appendix I Graph 12) Although the phenomena may have been a function of the methods with which the tests were scaled, the unique nature of the knowledge-application relationship, that of the knowledge retrieval being a prerequisite subskill for the more complex application items, is a possible explanation. A student would logically not be able to apply any more knowledge than he could recall, thus the application scores would necessarily be relatively lower than the knowledge scores. This artifact does not affect the interpretation of the product moment correlations but the lack of heteroscedasticity might influence some covariance measures, discussed later.

Multivariate Test   As an exploratory device, the multivariate test is analogous to using the overall F-test to decide whether it is appropriate to make individual comparisons. Therefore, a significant multivariate W allows the experimenter to continue his analysis by making univariate F-ratio comparisons. A significant F-ratio then enables one to calculate multiple and individual comparisons.

Both multivariate normality and homogeneity of variances and covariances were assumed. The tests for both are quite complex. Observation of the score distribution, and the sums of squares and cross products demonstrated no irregularities. (See Tables 4.5 and 4.6)

Table 4.7 shows that the prior division into achievement levels accounted for overwhelming differences. (χ^2 = 323.75 for 4 df, $p < .001$)

TABLE 4.5

MEANS AND STANDARD DEVIATIONS FOR THE KNOWLEDGE AND APPLICATION SCORES

ACHIEVEMENT LEVEL GROUPS n=16	KNOWLEDGE		APPLICATION	
	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION
Upper Third				
I - 1	23.06	4.29	19.68	3.89
2	22.68	4.09	16.37	5.45
3	22.18	4.94	18.75	5.56
4	23.43	5.09	19.81	6.35
5	22.25	3.38	17.18	5.84
6	18.68	4.24	16.50	8.03
7	20.93	5.27	18.75	7.12
8	21.43	4.05	19.18	6.33
Middle Third				
II - 1	18.25	3.61	11.31	6.79
2	19.81	3.53	14.87	5.30
3	18.00	3.35	12.68	6.05
4	18.37	3.61	11.62	5.61
5	16.37	5.51	9.75	5.53
6	14.06	5.64	11.43	5.88
7	12.81	5.54	8.00	5.30
8	14.25	5.07	13.56	5.38
Lower Third				
III - 1	10.43	4.50	7.62	5.15
2	12.00	4.52	5.56	4.40
3	10.81	3.76	5.68	4.27
4	12.00	4.52	7.06	3.89
5	10.00	5.15	7.06	6.16
6	8.87	4.18	4.06	3.86
7	9.00	3.41	3.87	4.24
8	7.88	4.73	5.37	4.48

1 = Variation I with feedback
2 = Variation I without feedback
3 = Variation II with feedback
4 = Variation II without feedback
5 = Variation III with feedback
6 = Variation III without feedback
7 = Variation IV with feedback
8 = Variation IV without feedback

TABLE 4.6
SUMS, SUMS OF SQUARES AND CROSS PRODUCTS OF KNOWLEDGE (X) AND APPLICATION (Y) SCORES

GROUP n=16	EX	EX^2	EY	EY^2	EXY
Upper Third					
I-I	369	8,805	315	6,429	7,449
2	363	8,487	262	4,736	6,113
3	355	8,243	300	6,088	6,910
4	375	9,177	317	6,885	7,611
5	356	8,092	275	5,239	6,274
6	299	5,857	264	4,968	5,044
7	335	7,431	300	6,386	6,692
8	343	7,599	307	6,491	6,899
	2,795	63,691	2,340	47,222	52,994
Middle Third					
II-I	281	5,131	181	2,739	3,402
2	317	6,467	238	3,962	4,808
3	288	5,352	203	3,125	3,720
4	294	5,602	186	2,634	3,506
5	262	4,746	156	1,980	2,787
6	225	3,641	183	2,611	2,929
7	205	3,087	128	1,446	1,880
8	228	3,634	217	3,377	3,417
	2,100	37,560	1,492	21,874	26,449
Lower Third					
III-I	167	2,067	122	1,328	1,407
2	192	2,610	89	785	1,306
3	173	2,083	91	791	1,061
4	192	2,610	113	1,025	1,573
5	160	1,998	113	1,367	1,408
6	142	1,522	65	487	780
7	144	1,470	62	510	678
8	126	1,328	86	764	841
	1,296	15,688	741	7,057	9,054
Grand Totals	6,191	117,039	4,573	76,153	88,497

1 = Variation I with feedback
2 = Variation I without feedback
3 = Variation II with feedback
4 = Variation II without feedback
5 = Variation III with feedback
6 = Variation III without feedback
7 = Variation IV with feedback
8 = Variation IV without feedback

TABLE 4.7

THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE MULTIVARIATE ANALYSIS OF THE COMBINED KNOWLEDGE AND APPLICATION SCORES (N = 384)

Source of Variation	W	p(k-1) df	χ^2*	P
Between Means	.36	2(24-1) 46	378.03	.001
Achievement Levels	.423	2(3-1) 4	323.75	.001
Step Size	.895	2(4-1) 6	45.14	.001
Feedback	.99	2(2-1) 2	.57	.05
Feedback x Step Size	.952	2(4-1)(2-1) 6	19.61	.01
Achievement Level x Step Size	.974	2(3-1)(4-1) 12	9.62	.05
Achievement Level x Feedback	.979	2(3-1)(2-1) 4	7.77	.05
Feedback x Step Size x Achievement Level	.99	2(4-1)(2-1)(3-1) 12	.57	.05

*-n log W as χ^2 is only an approximation to W distribution

The combination of entry repertoire, intelligence, and reading ability that makes up science achievement in this case, accounted for the major portion, (approximately 90 per cent) of the between-means variation. The feedback x step size interaction also produced significant differences ($\chi^2 = 19.61$ for 6 d.f., $p < .01$), thus allowing us to accept the statement that: providing knowledge of correct response has differential value for programs of varying step size when considering both knowledge and application test scores, in combination, as a criterion.

The main effects, step size and feedback, with their interaction significant, cannot justifiably be considered as independent sources of variation. It is interesting to contrast, however, the overall significance of the step size variable and the negligible overall amount of dispersion attributable to feedback. Not obtaining even chance difference due to feedback would in some instances be a sign of mismanagement of variable control. Having arbitrarily taken the extremes, with and without, as examples of the possible feedback dimensions, dispels any suspicion regarding experimental variable manipulation.

Appropriate Univariate Tests

As was mentioned earlier, the finding of significant multivariate W's due to achievement level and the feedback x step size interaction allows the experimenter to continue with their corresponding univariate F-tests. The complete three-way factorial breakdown of the sums of square s for both knowledge and application scores, available in Tables 4.8 and 4.9, offers more clues to the distribution

TABLE 4.8

THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE KNOWLEDGE SCORES (N=384)

Scource of Variation	Sum of Squares	d.f.	Mean Square	F
Achievement Levels	8,792.82	2	4,396.41	218.40 ***
Step Size	783.61	3	261.60	12.99 **
Feedback	0.00	1	0.00	
Feedback x Step Size	185.03	3	61.67	3.06*
Achievement Levels x Step Size	129.47	6	21.57	1.07
Achievement Levels x Feedback	16.19	2	8.09	.40
Achievement Levels x Feedback x Step Size	67.94	6	11.32	.56
Within Cells	7,250.19	360	20.13	

***$p < .001$ **$p < .01$ *$p < .05$

TABLE 4.9

THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE APPLICATION SCORES (N=384)

Scource of Variation	Sum of Squares	d.f.	Mean Square	F
Achievement Levels	9,999.75	2	4,999.87	167.49 ***
Step Size	187.52	3	62.50	2.09
Feedback	9.69	1	9.69	.29
Feedback x Step Size	164.78	3	54.92	1.83
Achievement Levels x Step Size	139.19	6	23.19	.77
Achievement Levels x Feedback	202.51	2	101.25	3.39#
Achievement Levels x Feedback x Step Size	244.18	6	40.69	1.36
Within Cells	10,746.19	360	29.85	

*** $p < .001$ # did not reach sig on overall test

TABLE 4.10

MAIN EFFECT AND INTERACTION MEANS FOR KNOWLEDGE AND APPLICATION TESTS

ACHIEVEMENT LEVELS	UT	21.83	(UPPER THIRD)		UT	18.28		
	MT	16.40	(MIDDLE THIRD)		MT	11.65		
	LT	10.12	(LOWER THIRD)		LT	5.78		
STEP SIZE	I	17.59	(PROMPTED VERSION)		I	12.57		
	II	17.46	(BASIC VERSION)		II	12.60		
	III	15.04	(MODERATE ")		III	11.00		
	IV	14.38	(DIFFICULT ")		IV	11.45		
FEEDBACK	w	16.11	(WITH)		w	11.75		
	w/o	16.12	(WITHOUT)		w/o	12.06		
		w	w/o			w	w/o	
FEEDBACK ×	I	17.00	18.14		III	12.85	12.25	
STEP SIZE		17.00	17.91			12.37	12.81	
	II	16.18	13.85		IV	11.31	10.65	
		14.23	14.52			10.19	12.70	
		UT	MT	LT		UT	MT	LT
STEP SIZE ×	I	23.86	18.84	11.31	I	18.00	13.07	6.58
ACHIEVEMENT LEVELS	II	23.00	18.33	11.50	II	19.25	12.15	6.36
	III	20.63	15.34	9.51	III	16.82	10.58	5.55
	IV	21.18	13.64	8.50	IV	18.96	10.76	4.62
		w	w/o			w	w/o	
FEEDBACK ×	UT	22.07	21.56		UT	18.56	17.96	
ACHIEVEMENT LEVELS	MT	16.16	16.60		MT	10.43	12.85	
	LT	10.05	10.17		LT	6.05	5.51	

of variance, although only the single main effect and single interaction can be rightfully considered. Table 4.10 exhibits the main effect and interaction means. Graphs of the three two-way interactions are found in Appendix I, #1 and #2.

Entry achievement level had almost equally strong an effect on the variation of both types of criterion behaviors. (Knowledge F = 218.4 at 2 and 360 d.f., p .001 and Application F = 167.49 at 2 and 360 d.f., p .001). Even when the variance common to both knowledge and application scores is covaried out, a significant amount of variance still remains. (F = 12.09 at 2 and 359 d.f., p .001) (See Table 4.11).

TABLE 4.11

ANALYSIS OF THE APPLICATION SCORE VARIANCE DUE TO THE ACHIEVEMENT LEVEL MAIN EFFECT THAT REMAINS WHEN KNOWLEDGE SCORES ARE COVARIED OUT

Source of Variation	Sum of Squares	d. f.	Mean Square	F
Achievement Levels	504.81	2	252.41	12.09*
Residual	7,513.66	360	20.87	

$b_{res} = \frac{4,841.29}{7,250.19}$ $b_{tot} = \frac{14,290.41}{16,043.01}$

$b_{res} = .6677$ $b_{tot} = .8907$

* p .001

It is interesting to note that effects of the step size variable alone, even though its interaction with feedback is more important in terms of the research hypotheses. The dispersion needed to carry the overall test beyond significance is situated primarily within the knowledge measures. The same covariance technique as was used with the achievement level above reduces the ratio of differences due to step size within the application measures to .99. (F = 0.99 for 3 and 359 d.f., $p > .05$) (See Table 4.12).

TABLE 4.12

ANALYSIS OF THE APPLICATION SCORE VARIANCE DUE TO THE STEP SIZE MAIN EFFECT THAT REMAINS WHEN THE KNOWLEDGE SCORES ARE COVARIED OUT

Source of Variation	Sum of Squares	d. f.	Mean Square	F
Step Size	61.89	3	20.63	0.99*
Residual	7,513.66	360	20.87	

$b_{res} = \frac{4, 841.29}{7, 250.19}$ $b_{tot} = \frac{5,194.37}{8,033.80}$

bres = .6677 $b_{tot} = .6465$

* p .05

The feedback x step size interaction reaches the .05 level of significance for the knowledge measure (F = 3.06 for 3 and 360 d.f., p .05) but not for the application measure (F = 1.83 for 3 and 360 d.f., p .05). When the knowledge scores are covaried out, the ratio increases beyond the .05 level (F = 2.87 for 3 and 359 d.f., p .05) (See Table 4.13). The combining effects of step size and feedback had an effect upon application that was independent from that which it demonstrated on the knowledge test.

TABLE 4.13

ANALYSIS OF THE APPLICATION SCORE VARIANCE DUE TO THE STEP SIZE X FEEDBACK INTERACTION WHEN THE KNOWLEDGE SCORES ARE COVARIED OUT

Source of Variation	Sum of Squares	d. f.	Mean Square	F
Step Size X Feedback	179.89	3	59.86	2.87*
Residual	7,513.66	360	20.87	
b_{res} = .6677		b_{tot}	= .6578	

*p .05

The reason for the unexpected increase in the step size x feedback interaction variance while the remaining variance of both main effects decreased can be found in their respective between mean correlations (See Table 4.14).

The residual covariance accounted for or removed 36% of the application error variance, the denominator of the F - ratio. Both main effects, achievement level and step size, removed 99% and 92% respectively from the "true" numerator variance. The overall effect was to decrease markedly the ratio. The step size x feedback interaction "true" covariance between knowledge and application scores was small, accounting for only 8%. Removing 36% from the denominator error application variance and only 8% from the numerator "true" variance had the effect of raising the F - ratio above its non-covaried estimate, from 1.83 to 2.87. The increased ratio was significant at the five per cent level.

TABLE 4.14

BETWEEN MEAN AND RESIDUAL CORRELATIONS BETWEEN KNOWLEDGE AND APPLICATION SCORES FOR THE SIGNIFICANT SOURCES OF VARIATION

	Achievement Levels	Step Size	F x SS Interaction
Between	.995 (.99)*	.92 (.85)	.29 (.08)
Residual	.55 (.36)	.55 (.36)	.55 (.36)

*( ) denotes r^2

Another way seeing the covariance effects is to compare the step size x feedback interaction plots on Graphs #1 and #2 of Appendix I. The two plots, when contrasted, accentuate the figural differences. The significant rise within Variation IV on the application plot accounts for the major disparity.

Knowledge Criterion Hypotheses

Hypothesis #1 stated:

> In terms of the knowledge scores:
> Providing knowledge of correct response does not increase the effectiveness of small step size programs for students at all achievement levels.

It was expected that there would be a small but non-significant difference between feedback versions of the first, prompted variation. A larger, but still non-significant, difference between the with and without versions of the basic program was anticipated. Only one group of the six possibilities (three levels for each variation), the upper third studying the easiest program with feedback, had a higher mean score than its without feedback counterpart (See Table 4.15). This

TABLE 4.15

KNOWLEDGE SCORE MEANS FOR EACH OVERALL ACHIEVEMENT LEVEL AND TREATMENT GROUP (N=16 PER CELL)

Achievement Levels	Step Size Variation	I		II		III		IV	
	Feedback Version	W	W/O	W	W/O	W	W/O	W	W/O
	Upper Third	23.06	22.68	22.18	23.43	22.25	18.68	20.93	21.43
	Middle Third	18.25	19.81	18.00	18.37	16.37	14.06	12.81	14.25
	Lower Third	10.43	12.00	10.81	12.00	10.00	8.87	9.00	7.88
Overall Means		17.24	18.16	17.00	17.93	16.21	13.87	14.25	14.52

difference, .38 of a point, was clearly not significant. A difference of 2.58 was needed to reach the 5 per cent level. Clearly, providing feedback to the first two step size variations did not increase the ability of students, from any prior achievement level, to recall the program facts and concepts.

Hypothesis #2 stated:

> In terms of the knowledge scores:
> Providing knowledge of correct response does increase the effectiveness of moderate step size programs at all achievement levels.

Variation III was considered the moderate step size variation. The students took, on the average, twice as many seconds to respond to each frame as they did on the easiest program. The accuracy of the error rate data was so hindered by the suspected response-peeking, that no reliable estimate of the more common characteristic could be made.

Table 4.15 demonstrates the cell and overall means showing a marked reversal of the first two variations where the versions without feedback produced higher scores. All three level mean differences between with and without feedback versions favor the with feedback version. Only the upper third achievement level group difference, 3.57, was greater at the 5 per cent level. The other two levels had mean differences in the expected direction, middle third of 2.31, and lower third of 1.13. The over achievement level mean difference of 2.34 was also significant ($t=2.55$ for 154 d.f., $p<.01$). The needed difference for significance in this case was considerably smaller due, of course, to the tripling of the sample size by adding the scores of three levels together.

The decision is that the multiple hypothesis is partially correct. The overall and upper third significances support while the middle and lower third differences did not reach the needed difference and, therefore, do not support the hypothesis.

Hypothesis #3 stated:

> In terms of the knowledge scores:
> Providing moderately difficult frames with knowledge of correct response will be more effective than any other combination of step size and feedback for all achievement levels.

Table 4.15 clearly demonstrates that the above hypothesis is rejected at all achievement levels. None of the Variation III means are larger than either of the first two variations. The smaller step size variations means differences are, however, not large enough, 2.87, to be considered significantly more different.

The third or moderate variation was, however, more efficient as the students took on the average, 11 minutes less to work through the program. The moderately difficult program with feedback was not more effective than the other versions, but was possibly more efficient.

Application Criterion Question

With no research on which to establish hypotheses regarding the application measures, only after-the-fact data analysis was acceptable.

> What effect will programs of varying step size and feedback have upon the learner's ability to apply the concepts within these programs to similar situations?

To recapitulate the earlier findings: 1) Achievement level was still an effect after the knowledge scores had been covaried out;

2) Neither step size or feedback was an independent influence; 3) The step size x feedback interaction became significant after the knowledge score interaction was covaried out.

The major disparity between interaction plots was demonstrated between the two versions of the most difficult Variation IV (See Table 4.16). The group using Variation IV, without feedback, scored 2.49 points higher than did the group with feedback. The without group, it may be remembered, worked an average of 20 seconds more per response. Those students studying Variation IV, without feedback, also scored as well as those studying from the easiest variation with feedback.

Retention Loss Question

> What effect will programs of varying step size and feedback have upon the learner's ability to retain the program concepts over a long interval of time?

The retention measures have been recorded as difference scores, first knowledge score minus second knowledge score for each student. The resulting difference measures represent the amount of information loss over the summer months. Complete data were available for the knowledge test only. Six points were added to every difference score to eliminate all negative values. The three-way factorial univariate analysis of variance resulted in only the achievement level main effect reaching significance, (F = 4.46 for 2,158 d.f., p .05) (See Table 4.17). Limited returns from the application test tended to show approximately the same result, general loss over all treatment combinations.

TABLE 4.16

APPLICATION SCORE MEANS FOR EACH OVERALL ACHIEVEMENT LEVEL AND TREATMENT GROUP
(N=16 PER CELL)

Achievement Levels								
Step Size Variation	I		II		III		IV	
Feedback Version	W	W/O	W	W/O	W	W/O	W	W/O
Upper Third	19.68	16.37	18.75	19.81	17.18	16.50	18.75	19.18
Middle Third	11.31	14.87	12.68	11.62	9.75	11.43	8.00	13.56
Lower Third	7.62	5.56	5.68	7.06	7.06	4.06	3.87	5.37
Overall Means	12.87	12.27	12.70	11.30	11.30	10.66	10.21	12.70

TABLE 4.17

THREE-WAY FACTORIAL SOURCE OF VARIATION TABLE FOR THE RETENTION LOSS SCORES (N = 192)

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Achievement Levels	124.54	2	62.27	4.78*
Feedback	23.38	1	23.38	1.79
Step Size	59.89	3	19.96	1.53
Achievement Levels x Feedback	3.04	2	1.52	
Feedback x Step Size	48.43	3	16.14	1.24
Achievement Levels x Step Size	23.13	6	3.85	
Achievement Levels x Feedback x Step Size	36.46	6	6.07	
Within Cells	2,203.13	169	13.04	

* p .05

Affective Data

The interpretation of affective scores presents somewhat of an enigma. Independent analyses of the adjective rating scales would be statistically incorrect, for one violates the same independence assumptions that called for multivariate measures in the cognitive tests.

The only solution was to calculate the independent variance analyses, compute the rating intercorrelations, and interpret by combining the two as the multivariate would if it were available. Therefore, the reported probabilities from one scale are not independent of the probabilities of the other scales. (See Table 4.18 and related Graphs #5 through #11 in Appendix I.) The intercorrelations between ratings ranged from -.04 to +.74, with the majority centering around +.40. Some combinations were more highly correlated with the others, as can be seen from the table of probabilities

(See Table 4.18). Difficulty level, progress - no progress, successful - frustrated, clustered together on a "difficulty" dimension, while rewarded - punished and alert - careless were related to the main group but in somewhat different ways. Interested - bored was the most independent of all the ratings.

TABLE 4.18

COMPILATION OF THE F-TABLE PROBABILITIES FOR THE SIGNIFICANT MAIN EFFECTS AND INTERACTIONS ON THE AFFECTIVE RATING DATA

	Achievement Level	Feedback	Step Size	Achievement Level x Step Size	Step size x Feedback	Achievement Level x Feedback	Achievement Level x Feedback x Step Size
Difficulty Level	.02	---	.001	---	.025	---	---
Alert - Careless	---	---	.05	---	---	.01	---
Rewarded - Punished	.03	---	---	---	---	---	---
Progress - No Progress	---	---	.005	---	---	---	---
Successful - Frustrated	.001	---	.001	---	---	---	---
Interested - Bored	---	.03	---	---	---	---	---
Average Affect	---	---	.001	---	---	---	---

The interested - Bored continuum, selected to represent the "pall" effect, was significantly affected only by the giving and withholding of feedback. Not knowing whether they were correct or incorrect proved to be more boring for the students. Neither achievement level or step size produced a noticeable variation in the interest-bored ratings.

Hypothesis #4 states:

> In terms of the knowledge scores:
>
> Providing moderately difficult frames with knowledge of correct response will reduce the boredom or "pall" effect among the upper and middle third achievement levels.

Table 4.19 demonstrates the mean values for the upper and middle third groups taking the with feedback versions. The upper third group means show a slight rise over variations, meaning that each group felt more bored than the last. This obviously does not agree with the hypothesis. The middle third group means, however, do exhibit the hypothesized trend, with the moderately difficult Variation III being the most interesting. The differences did not, however, reach the 2.03 needed for the .05 level of significance. The averages of the two levels also follow the hypothesized trend.

In regard to question #2:

> Will there be any differences in the frequency of boredom as reported by the below average achievers who are using the various experimental programs?

The lower group interested - bored rating averages, demonstrated considerable variation (See Table 4.20). Those students who received feedback generally exhibited more interest in the programs. The group studying with step size Variation II, without feedback, however, rated themselves as most interested.

Cognitive, Affective and Cognitive-Affective Correlations

As an exploratory device, separate Pearson product-moment correlations were computed for all meaningful measures using students taking the eight different program versions as distinct groups. This resulted in eight intercorrelation matrices whose sample sizes ranged from 38 to 48. As would be expected from the large entry achievement level main effect differences, the correlations between entry achievement scores and both knowledge and application scores remained high, ranging from + .62 to + .83 (See Table 4.21).

The correlations between the affective ratings ranged from -.04 and +.74. When averaged over the eight groups, the correlations were all in the vicinity of +.40. A Pearson product moment r greater than .28 would be considered significantly larger than .00 at the .05 level.

The correlations between both knowledge and application scores, and the affective ratings ranged from +.03 to -.55. The great majority, however, remained around -.10, denoting little relationship between the cognitive measures and the affective ratings.

TABLE 4.19

COMPARISON OF THE INTERESTED - BORED RATING MEANS FOR THE UPPER AND MIDDLE THIRD ACHIEVEMENT LEVEL GROUPS USING THE WITH FEEDBACK VERSIONS (N=14 PER CELL)

Step Size Variations -	I W/Feedback	II W/Feedback	III W/Feedback	IV W/Feedback
Achievement Levels:				
Upper Third	3.08*	3.64	3.70	3.85
Middle Third	4.21	3.50	3.42	4.14
Overall Averages	3.64	3.57	3.56	3.99

* Higher scores represents more boredom, low scores more interest

TABLE 4.20

COMPARISON OF THE INTERESTED - BORED RATING MEANS FOR THE LOWER THIRD ACHIEVEMENT LEVEL GROUPS USING ALL VERSIONS (N=14 PER CELL)

Step Size Variations	I	II	III	IV
Lower Third Achievement Level Versions:				
With Feedback	3.28	3.35	4.07	3.85
Without Feedback	4.28	2.28	4.35	7.42
Overall Averages	3.78	2.82	4.21	5.64

TABLE 4.21

CORRELATIONS BETWEEN ENTRY ACHIEVEMENT AND KNOWLEDGE AND APPLICATION SCORES FOR THE EIGHT PROGRAM GROUPS

Step Size Variation	I		II		III		IV	
Feedback Version	W	W/O	W	W/O	W	W/O	W	W/O
Entry Achievement with Knowledge	.80	.81	.71	.74	.75	.69	.79	.83
Entry Achievement with Application	.70	.62	.73	.73	.71	.73	.76	.71

Results Summary'

In terms of theoretical hypotheses, the study fared rather well. Data confirmed the initial theoretical position (Hyp. #1) that immediate feedback would not increase the effectiveness of the small step size (mean error rate below 10%) programs at any achievement level, when considering knowledge and comprehension items, as criterion. Conceptual meanings were formed and factual associations were made within the highly redundant context of the frames alone, without the presence of the correct answer.

By decreasing both intra- and inter-frame redundancy, it was further shown that the focus of these associations could at times be shifted from the frames such that feedback now became informative (Hyp. #2). This phenomenon occurred over all levels although the middle and lower third increase did not quite reach significance at the .05 level. As might be expected, the higher achievers were able to gain and retain more from the now informative feedback, as the differences clearly decreased in line with achievement levels.

The groups taking the "moderate" difficulty level program with feedback did not, as hypothesized (Hyp. #3), demonstrate increased knowledge scores over groups using the other versions. Both groups given the with and without feedback versions of the smaller step size programs produced higher scores. If one, however, considered efficiency rather than effectiveness, the moderate version with feedback took, on the average, about one-half as much time to complete as did the smaller step-size versions.

Both prior achievement and program knowledge scores were significantly related to application scores. (Question #4) Neither step-size or feedback variations had any effect upon application, but there was a significant interaction of the two when the knowledge score interaction was covaried out. (Question #2)

Only prior achievement level accounted for a significant amount of variance in the knowledge retention scores. (Question #3) The lower third lost less than the other achievement level groups. The phenomenon could be explained as resulting from a combination of regression and "floor" effects, in the sense that their scores were near zero.

Both error rate and time to complete the program were affected in the expected ways by all the main variables and their interactions.

There was a trend toward less boredom in the moderate with feedback version among the middle, but not the upper third achievement level groups. (Hyp. #4). In relation to research Question #5, the lower third group rated themselves, on the average, as most interested in Variation II.

Step-size variation influenced the affective ratings related to success and difficulty. Feedback variation accounted for the only significant variation in the interested-bored ratings. None of the affective ratings on the average correlated with the knowledge or application cognitive measures. See Table 4.22 for a complete summary.

Discussion Overview

The overall plan for the discussion section will be to: 1) attempt to answer the questions posed in the problems section; 2) identify and clarify new problems that developed within the investigation and attempt some rationale for their happening; 3) offer the proposed contingent generalizations growing out of the study findings and relate them to the results of other studies; and 4) draw up some tentative

Table 4.22

COMPILATION OF THE F-TABLE PROBABILITIES
FOR THE SIGNIFICANT MAIN EFFECTS AND
INTERACTIONS ON MAJOR CRITERION MEASURES

	Step Size	Feedback	Step Size x Feedback	Achievement Level	Step Size x Achievement Level	Feedback x Achievement Level	Step Size x Feedback x Achievement Level
Major Criterion Measures							
Cognitive:							
Error Rate	.001	.001	.001	.001	.001	.001	.01
Time to Completion	.001	.001	.01	.001	.01	.05	--
Knowledge Post Test	.01*	--	.05	.001	--	--	--
Knowledge Retention Test	--	--	--	.05	--	--	--
Application Post Test	--	--	.05**	.001	--	--	--
Affective:							
Difficulty Level	.001	--	.025	.02	--	--	--
Alert-Careless	.05	--	--	--	--	.01	--
Rewarded-Punished	--	--	--	.03	--	--	--
Progress- No Progress	.001	--	--	.001	--	--	--
Interested-Bored	--	.03	--	--	--	--	--
Average Affect	.001	--	--	.08	--	--	--

* Not independent of SS x F interaction

** Significant when Knowledge covaried out guidelines for the strategy of programing based upon these contingent generalizations.

Discussion of Problems Section Questions

> Is it, as the linear programed instruction originators stress; most effective and efficient to provide a highly redundant discourse such that there is a high probability that the student will respond correctly, and then follow immediately with the correct response? Or should one provide less redundant narrative such that the student can only offer a tentative response and then give him the correct answer?

It seems that the present criterion of less than ten percent error is too demanding. The study demonstrated that the prior to response information can be less redundant, thus more difficult, without causing a major decrease in effectiveness. The less redundant frames individually tend to increase the individual frame response latency, but removal of others decreases the total time taken to complete the program. Making the items too difficult, to where the error rates exceeded approximately 40 per cent, seemed to be the point at which both effectiveness and efficiency seem to decrease. Some intermediate between 15 and 30 per cent error appear to be optimal.

There seems to be a functional relationship similar to that expressed by a negatively accelerated curve where more redundancy increases learning to a point at which it becomes necessary to question the amount of increase in retention in terms of the amount of time to repeat the information. The curves representing the step size x achievement level interaction in Appendix I, Graph I, give some evidence of this relationship. The curves in the chart have to be reversed along the step-size dimension to acquire the negative acceleration. The slopes of the curves are accentuated by the arbitrary differentiation along the knowledge score dimension.

> Would one strategy be more dependent upon the information within the correct ånswer than the other?

The hypothesized interaction occurred, showing feedback to have a facilitating effect only at about the same error rate range as above, 15 to 30 per cent. Both more and less redundant frames seemed to function equally as well with or without feedback.

> Are there different strategies of information introduction for particular classes of learners?

Step-size variation did not exhibit differential influence upon any of the three achievement levels. There were no step-size achievement level interactions within any of the cognitive criterion variance analyses.

Ausubel's hypothesis that "One might expect feedback to be more effective in the case of the less able students" (p. 205) proved to be both correct and incorrect. It was correct in that the reinforcement aspect of feedback did generate greater interest, but incorrect in that there was no concomitant increase in the cognitive outcomes. This inability of feedback to exhibit a significant differential in cognitive outcomes, within the lower third achievers, may have been masked by their lack of developmental readiness for abstract concepts and principles, and the fact that they were conveyed in symbolic form. In this light, the hypothesis as to the facilitative effect of feedback upon cognitive outcomes for low achievement students remains feasible under certain conditions.

> Are there concomitant affective states which accompany these information strategies that tend to enhance or interfere with the cognitive functioning?

There was an indication of increased interest or reduced boredom when programs were moderately difficult, especially within the middle third achievement level. The lack of positive correlation, however, between any affective rating and either knowledge or application score, makes the probability of affective change affecting cognitive functioning somewhat doubtful, at least for a task of this duration.

> Is there a single strategy using the linear programed instruction framework that is most effective in attaining both application and long-term retention objectives?

Since no concerted effort was made to teach application skills, the lack of differences from the major variables might have been expected. The findings, however, are interesting in that gross manipulations of the major variables had so little an effect. The high correlation between the application socres and prior achievement test scores and the added explained variance from the program knowledge scores, offer some leads for programers and researchers. These leads are discussed in the section about cognitive variables.

The absence of any reliable effects upon delayed retention loss over a span of five months was not surprising. One really could not expect even quite drastic variation in the programs to influence long-term retention with none of the known facilitators such as spaced review added to the instructional strategy. Nevertheless, the students who scored high on the initial test maintained their superiority in the latter exam, so there is reason to ensure the maximum initial learning when no review is to be carried out. Whether there would be a general leveling effect with review remains to be tested.

> Finally, is there a single position, such as Skinner's, that can provide a rationale for generating solutions to the foregoing questions?

Hilgard has stated that he felt that none of three major theoretical positions, that of Skinner, Guthrie, and the cognitive theorists, could account for the major advances in programed instruction. The sections that follow provide an assessment of this study's findings relevant to these three positions in order to seek out either an agreement with Hilgard's opinion, or some variation which would be explanatory.

Operant Conditioning Position

The position that is being attacked is based upon Skinner's operant conditioning; this position will be discussed first. The only finding within the present study that significantly demonstrated the influence of "small step-size" or "immediate reinforcement" was the greater interest shown by those groups who received feedback. It seems that such behaviors as "paying attention" and "keeping at the task", which would be akin to "being interested", were facilitated by giving the correct answer. The use of "immediate reinforcement" may then, be more appropriate for the control and maintenance of the instructional situation rather than for any learning objectives, as such. The two objectives are obviously not independent, however, as staying with the task would eventually increase the total amount of learning.

Contiguity Position

The contiguity position advocated by Guthrie is best demonstrated by the work of one of his pupils, Lumsdaine. The present study further substantiated the original results of Angell and Lumsdaine (1959) in a paired-associate learning task, later supported by Wittrock and Twelker(1964) in a problem-solving task consisting of ten enciphered sentences. To compare more fully the related results, the original hypotheses and findings of the three studies will be repeated.

Angell and Lumsdaine's original hypothesis in the paired-associate task read, "To the extent that adequate prompting was provided for overt response, the efficacy of (and need for) the further implicit-response prompting provided for by the confirmation panel would be expected to diminish." (p. 478). They found a significant interaction between prompting trials and correction treatment which resulted in two conclusions:

> "(1) A prompting trial (where both stimulus and response words were shown together) was more effective than an anticipation trial (separation of stimulus and response terms) when partial correction was employed but not when full correction was used,
>
> and
>
> (2) Full correction was superior to partial correction when there had been no initial prompting trials but not when there had been more than two trials."

This final clause provided the tie-in, for when there was a redundant display, two or more prompting trials, there was no facilitating effect from increased feedback.

Wittrock and Twelker stated their belief in another way as they hypothesized that "knowledge of correct response enhances learning, retention, and transfer when the information it contains is not greatly redundant." They found that "non-redundant information in the form of knowledge of correct response added to a minimally directed situation enhanced learning while redundant knowledge of correct response added to an already prompted situation did little to learning, retention, and transfer." (1964, p. 17).

The relation to the present study's rationale and findings can best be made by combining the present study's first two hypotheses:

> In terms of knowledge recall, providing knowledge of correct response does not increase the effectiveness of small step-size programs but does increase the effectiveness of moderate step-size programs at all achievement levels.

Students performed equally as well when studying from highly redundant frames without feedback as they did with feedback. Feedback facilitated criterion performance only when the redundant information within and between frames was reduced to where it took, on the average, twice as long as the easiest program to respond to a frame, and resulted in a mean error rate of about 20 per cent. In addition, feedback became inoperative again as further redundant information was removed until the mean error rates approximated 50 per cent or greater.

The commonality among the three positions and tasks was that all the information needed to make the correct overt response was available prior to that response, thus making feedback or reinforcement unnecessary. As soon as parts of that information were removed or spread temporally apart, the overt response became less available and made the inherent information within the feedback necessary. The characteristics of the prior information are that it be: (1) temporally close together, specifically in the case of the paired associate task, (2) self-explanatory or directive, in the case of the problem-solving task, and (3) contextually redundant in the case of the programed material or subject matter comprehension task.

Therefore, the contiguity position holds in that paired associates can be related, problems can be solved, concepts may be learned, effectively and efficiently, from judiciously selected and placed information.

Cognitive Theorists' Position

It seems that the emphasis upon making an overt response, that it be correct, and that it be immediately reinforced, has been unwarranted. Evidence has been advanced to show that students perform somewhat independently of all three of these characteristics. Hilgard discredited the notion of the importance of the response in complex, meaningful learning by saying that "to talk of these little replies were what was being learned is to make a travesty of the educational process" (1964, p. 136). In doing so, he sided with the cognitive theorist

> "who is concerned with knowing and understanding, ... and is little likely to believe that the essence of programming is small steps, responding, and reinforcing; and would also say These are only mechanical details related to something far more significant that is going on" (p. 136).

That "something more significant that is going on" can only be surmised at this time. A number of variables of interest to the cognitive theorists, however, offer some possible explanations of that "something more significant." A discussion of some of these follows.

Cognitive Maturity Gotkin (1964) viewed the problem he faced in trying to program for "socially disadvantaged" seventh graders, much like the lower third achievement group of the present study, in terms of the developmental scheme proposed by Piaget and adapted by Bruner.

> "Bruner describes three levels by which the child represents his experiences: the most primitive is the enactive, which is based on motor activity; the second is the iconic, which is based upon images of what is perceived; the third is symbolic, in which experience is represented in either natural or formal language" (p. 5).

Gotkin's research team was "continually faced with the inability of children at the seventh grade level to cope with iconic, let alone symbolic representations" (p. 5).

The basic program and its variations in the present study was, except for a few simple drawings, wholly symbolic. The program content involved a verbal representation of enactive demonstrations. The demonstrations, in turn, were examples of the abstract concepts and principles of electricity. It undoubtedly would have been developmentally sound to represent the enactive demonstrations, i.e. rubbing a comb through hair, in an enactive, experiment-like form.

Bruner (1964) also forewarned of the difficulties in translating between these three representational forms. It would be unwise, therefore, to expect these students functioning at the concrete, enactive stage to be able to cope with criterion tests which were in the symbolic or most abstract form. The extension of this reasoning to the present study is that the lower third achievers were doubly handicapped, once by the symbolic program, and again by the symbolic criterion tests.

Entry Repetoire The role of entry repetoire was clearly the most important source of variation in the present study. Both Gagné (1965) and Ausubel (1963) have discussed different variations of this overall factor as being most influential in meaningful learning. Gagné has taken the position that to learn higher level concepts and principles one must have mastered the prerequisite lower order concepts and principles. Ausubel, on the other hand, cites the "availability in cognitive organization of relevant subsuming concepts at an appropriate level of inclusiveness" (p. 28) as crucial for incorporating new meaning. Both are obviously talking about a function within the prior achievement as defined and used as a levels factor in the present study.

If it were possible to analyze the overall learning task such that the prerequisite lower-order concepts and principles as well as the more inclusive, subsuming concepts were evident, then a test could be devised more specific than the overall test that was administered. This test could assess both the Gagné and Ausubel notions by predicting that certain learners, lacking either lower-order or more inclusive concepts, would be unable to incorporate the program concepts and principles into their repetoire. For example, the basic principle, unlike charges attract, could not be learned unless the three independent concepts were understood, a la Gagné. Ausubel might expect the more inclusive concepts of positive and negative (charges), magnetism, and electrons to be a part of the repertoire before comprehending the principle. In either case a more explicit, detailed look at entry repertoire might have served to elaborate upon the large prior achievement effect.

Individual Accommodation and Pacing There are essentially two rates at which one studies, his own and one regulated by another, such as the teacher. Programed instruction has been heralded for its characteristic of permitting self study or pacing. Unfortunately, the small step approach has been too slow a pace for many and resulted in the common boredom effect. The quick student resents having to wade through the redundancy both within and among frames. Programed instruction, therefore, only partially allows for self-pacing. Prose is much less restrictive as one can skim and does not have to overtly respond. There have been accounts of examples of fantastic reading rates with remarkable retention and comprehension. Oddly enough, textbooks have neither been praised for this inherent quality, nor have children been allowed to use this quality by reading ahead at their own pace.

Decreasing both intra- and inter-frame redundancy, as a way of manipulating step size, markedly effected the average pace at which students moved through the program. Removing redundancy from Variation III caused students to take, on the average, twice as many seconds, 42, to 21, to respond than Variation I. What kinds of activity dominated the extra 20 seconds? There was undoubtedly some rereading of the frame in the attempt to determine not only the desired response but the other word that was deleted. Could it be that the rereading of the frame was equivalent to the reading of the redundant frames that were removed?

Making the frames more difficult, that is, setting another pace, made everyone shift into another gear, so to speak. Some who could maintain that pace, were able to respond at the 20 second pace, took it in stride, while others increased their response latency until they either gave up or produced a response. Jacobs (1963) discussed this maximum point at which the response becomes apparent after much cognitive activity. He proposed a closer look at response latency rather than error rate. The same conclusion was drawn from the confidence rating data and now reappears in inspecting the data from the step-size variations.

Another interesting sidelight is that when feedback was available there was less time spent searching for a response. Some of this time was probably due to a better understanding, but it is hypothesized that availability of feedback reduced response latency because it was a tension release, much like peeking to find the murderer in a who-done-it.

The cycle of tension arousal, provisional try, and tension relief, hypothesized to occur at this level, it must be remembered, was self-paced. It is interesting to suppose what would have happened if not only the redundancy, or information timing, was manipulated, but also the amount of time to work out the frame response. This condition is much more similar to classroom interaction where the teacher can control both redundancy and time for response. What effect would these extra restrictions have upon the intra-frame cognitive functioning and related output?

Task Difficulty Bruner has proclaimed that anything can be taught at any level if it is "intellectually honest." And yet people agree that some concepts are more sophisticated than others. It is conceivable that the concept of electricity is such a sophisticated concept. If so, then it is also possible that regardless of the representive form of the teaching example, the average seventh grade student is not able to conceptualize the "sophisticated" subleties of induction, insulation, chemical action, etc., that were called for by the program. This failure would become more evident as the student attempted to apply these concepts and principles to new situations, as was noted.

Overall Conclusion as to Theoretical Rationale

The relation of the three theoretical positions to the results of programed instruction research seems to resemble the story of the three blind Indian fakirs trying to describe an elephant. Each fakir elaborated upon a separate part of the elephant, the trunk, tail and the hoof, and demanded his explanation to be correct. Obviously, each was partially accurate, as this investigator feels is also true in the case of the explanation of students learning from programed materials.

Deterline (1963) has argued that only two characteristics, analysis of objectives and repeated tryout, have emerged unscathed from the transfer of operant conditioning techniques to programed instruction. The study would add to these the reward value within "immediate reinforcement," as a method for maintaining interest and instructional control.

The notion of schedules of reinforcement cannot be dismissed as a technique for keeping students at the task of studying.

The fact that students can learn from the cueing within the frame without access to the correct response is in some measure a validation of contiguity theory. Facts, definitions, concepts, and principles depend upon temporal contiguity for their understanding. The principle, unlike charges attract, would be difficult to associate if not printed together. The effect of scrambling words upon understanding in prose materials is another example of the need for contiguous presentation of stimuli or words.

Cognitive theorists can claim the possible need for a given level of cognitive functioning before concepts can be formed from symbolic presentation. They also have pointed to the dependency upon the learning of prerequisite skills and concepts before new learning can take place. Another of their concerns is the regulation of the pace of information transmission in accordance with the individual style of the learner. Finally, the cognitive theorists consider the difficulty of the task, that is, the level of complexity of the concepts, stating that there might be a developmental readiness for concepts as well as their representational form.

In summary, it is believed that all three positions, that of Skinner, Guthrie, and the cognitive theorists, provide valid rationale for different facets of the extremely complicated task, that of learning subject matter content in the form of linear programs.

New Problems that Evolved

The finding that students could make very few errors on the program and then not retain and transfer this initial acquisition to the criterion tests is puzzling. Even those who were not given feedback made relatively few program errors, so it is not a case of response-peeking and copying. A possible solution lies in the method used for prompting the easiest variation, that of giving the first letter or two of the correct response. For example the lower third mean error rates for the first and second variations without feedback jumped from 23% to 64%. It could be that the students were guessing at the response from nearby responses without understanding why the response was correct. It seems that "letter-prompting" possibly is a deterrent in that it gives the learner another method for determining the correct response that is not facilitating the search activity that results in understanding and recall.

The problem of explaining the large prior achievement differences could have been avoided if the prior achievement test had included items testing both subconcepts and more inclusive "subsumers." The problem accentuated the need for more specific delineation of program content, prerequisites, and sequence than is usually the case.

Finally the problem became more apparent that the present style of programing needs revision or addition if long term retention and application are the professed objectives, as they most usually are. Long-term retention has its known facilitators which might not be expected to be a part of the program, per se. But application to new situations is an immediate goal and more ingenuity is needed to develop guidelines for its augmentation.

Contingent Generalizations

The search for contingent generalizations seemed most profitable when combining the findings of the present study and those of related research, primarily that of Krumboltz and Klaus. The following contingent generalizations are taken from this compilation.

1. If one begins with a program whose error rate is below 10% and adds redundant information either before or after the overt response, then:
    a. the error rate should decrease,
    b. boredom should increase, but
    c. there is likely to be no increase in recall, long-term retention, or transfer.
2. If one begins with a program whose error rate is below 10% and removes redundant information before or after the overt response or adds irrelevant information, then:
    a. the error rate should increase,
    b. frustration should increase while boredom should decrease, and
    c. there is likely to be no major change upon recall unless error rate increases to around 40%. Long-term retention and transfer will probably not be effected.
3. If one begins with a program whose error rate is below 10% and provides feedback, then:
    a. there will probably be no increase in recall until further redundant information is removed,
    b. there will probably be increased interest in the task.

Those generalizations involving student variables are:

1. if prior achievement is lacking, accomplishment of immediate retention or application will probably be minimal for complex tasks.
2. that neither high IQ or excellent reading ability will probably be able to overcome a lack in prior skill and knowledge attainment when attempting to accomplish program objectives.
3. that manipulating major program variables will probably not aid or hinder any student within different prior achievement levels.
4. that a student who does not recall the program content will probably be handicapped in further application tasks.
5. that negative affective states such as boredom and frustration probably do not have any influence upon student cognitive performance when the program takes less than two weeks to complete.

Conclusions Concerning the Role of Step Size and Feedback in Programing Strategy

Regardless of how the prior achievement or prerequisite knowledge and skills are somehow accounted for, there is still the question of the strategy of information introduction in linear programed instruction, i.e. relationship of step size and feedback. If feedback were to be maintained because of its saluatory effect upon interest in the task, then it seems plausible that a less redundant form of initial information would be the programing prescription. An error rate of 20 to 40 per cent might be desired for both effectiveness and efficiency. There is also the tendency that the interest or maximal stimulation level may be at its peak within this range.

There is still a possibility that increased cognitive activity as evidenced by the greater amount of time spent per frame might activate some of the skills that seem to be operative in applying concepts and principles to new situations. Again, much more work needs to be done to demonstrate the utility of the strategy of programed instruction in attaining higher level objectives.

Programers should, it seems, concern themselves less about the response characteristics, and more with the frame information so that it will be:

1. carefully analyzed to insure that prerequisite concepts be available or made available.
2. in the informational form at which the target audience is developmentally functioning.
3. not attempting to teach concepts which may be too abstract or sophisticated for the target audience.
4. not restricted to a less than ten per cent response error rate criterion and, therefore, more challenging.
5. more lengthy; that is, more information per response than is usually the case, especially as the program nears completion.

CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATIONS FOR RESEARCH

Background

What seemed so simple and clear in the early history of the development of programed instruction has now shown its true aspect - namely, that of an extremely complex phenomenon in meaningful verbal learning. The principles underlying linear programed instruction were developed by analogy from variables proven to be critical in operant conditioning. The validity of the analogy has not held. For example, the necessity for elicited response so essential in operant conditioning has been seriously questioned in programed instruction studies. Similarly, results as to the importance of small incremental tasks (step-size), and immediate reinforcement (feedback of the correct answer) have been contradictory. It has been suggested that to gain a better understanding of these phenomena, studies involving complex designs offering interactive information were needed. This study developed a three-way factorial design utilizing multiple dependent measures to study the hypothesized interaction between step size and feedback over differing achievement levels. A part of the study was devoted to assessing the adequacy of the student's subjective confidence in his frame response as a measure of step size.

This theoretical interest of the author and the practical interest of four 7th grade science teachers in the potential of programed instruction for aiding slower students combined to provide the necessary conditions for the conduct of the study.

Objectives

The study was designed to provide evidence concerning the following general questions:

1. Is the student's subjective confidence estimate of the certainty of the response a workable and meaningful operational definition for step size?
2. Does the knowledge of the correct response (feedback) have differential value for programs containing frames of varying step size? Are there separate interacting effects upon students who have differing entry achievement levels?
3. If so, is there an optimal combination of frame step size plus feedback such that a learner would be able best to comprehend concepts and principles and apply these principles to similar situations?
4. Is this combination optimally effective when considering students from varying levels of entry achievement?
5. What consequence, if any, will this optimal combination have upon the recurring boredom or "pall" effect? Will the consequences differ among students from varying levels of entry achievement?
6. What effect will programs of varying step size and feedback have upon the learner's ability to apply the concepts within these programs to similar situations or to retain the program content for long periods of time?

More specifically, the study tests the following hypotheses:

In terms of the comprehension of concepts:

1. Providing knowledge of correct response does not increase the effectiveness of small step size programs for students at all achievement levels.
2. Providing knowledge of correct response does increase the effectiveness of moderate step size programs at all achievement levels.
3. Providing moderately difficult frames with knowledge of correct response will be more effective than any other combination of step size and feedback for all achievement levels.
4. Providing moderately difficult frames with knowledge of correct response will reduce the boredom or "pall" effect among the upper and middle third achievement levels.

Procedures

A selected portion of a published program covering static electricity and voltaic cells was field tested and revised twice to reach the minimal error rate conditions needed. During this phase attempts were made to judge the usefulness of assessing the student's confidence in his frame response as an indicator of frame step size.

Three variations of the basic program were developed to serve as step size levels. The first contained one or two letter prompts for every response, thus representing the easiest version. Redundant and review frames plus key words within frames were systematically eliminated from the other two versions resulting in two more difficult step size levels. Each of the four variations was duplicated in a form containing the correct response and in one without to constitute the eight "treatment" materials.

At the same time a general science achievement test was being developed using differentiating items from previous teacher-made tests. The students were ranked according to their scores on this test, divided into thirds and randomly assigned to the eight program variations. The assignment procedure called for randomizing within each succeeding group of eight students while proceeding down the ranked list within the three "levels." The resultant was a 4 x 2 (step size x feedback) factorial experiment randomized within blocks design.

All available (approximately 400) 7th grade science students were assigned to programs. No sampling was undertaken. Two separate criterion tests were developed, one involving knowledge and comprehension items; the other, application items. A set of affective rating scales were logically developed and plans to take time estimates made.

The time to complete the phase of program and tests ranged from two to eight fifty-minute periods. Students finishing early were given remedial or enriching individual study.

Cell size was equated at 16 by random elimination for ease in statistical calculation. This operation brought the total to 24 x 16 or 384 subjects. Multivariate ratios were calculated for the pertinent cognitive sources of variance. Appropriate univariate F-tests, and individual comparisons were made to substantiate or refute the theoretical hypotheses and questions. Three-way factorial univariate analyses of variance were calculated for the affective scales, time estimates, error rates, and combined scores. Intercorrelation matrices were computed for all the possible affective and cognitive combinations for groups taking each of the eight program variations.

The knowledge and application tests were given again in the ensuing fall, approximately five months after the program administration. Only knowledge scores were meaningful as many of the application exams were not completed within the allotted school period time limit. A univariate factorial analysis of variance were computed on the knowledge loss scores.

Conclusions

Concerning the adequacy of the subjective confidence rating as a measure of step size:

1. The subjective confidence estimates were found to add little information to the dichotomous right-wrong of error rate.
2. Response latency seemed to be a more meaningful measure of step size.

Concerning the overall effects on knowledge and application scores:

1. Prior achievement level was a more important influence than was expected.
2. A small, in relation to achievement level effects, but significant interaction between step size and feedback was found. Taken as though they were independent, step size was, and feedback was not, a significant factor.
3. There were no step size or feedback by achievement level interactions.

Concerning the independent effects upon knowledge scores:

1. Prior achievement levels accounted for more variation than expected.
2. The step size × feedback interaction was significant allowing the test between treatment combinations which demonstrated:
    a. Feedback was not of value to the two small step size variations.

   b. Feedback became operative only within the framework of the moderately difficult version. The upper third was the only achievement level group to show significant gain over the without version groups. The middle and lower third level group differences were in the predicted directions, however, making the overall difference significant.

3. The effect of increasing step size (although not independent because of its interaction with feedback) significantly decreased the knowledge scores in the expected direction over all achievement levels.

Concerning the independant effects upon the application scores:

1. A significant variation remained due to prior achievement levels even when program knowledge-application score covariance was partialled out.
2. A significant step size x feedback interaction remained when the knowledge interaction was partialled out. This was due to what may or may not have been an artifact - a difference between the with and without feedback versions on the most difficult variation. The without version was the most effective, as effective as the easy variations with feedback. Further research is needed to determine the reliability and rationale underlying this finding.
3. Neither step size or feedback was an independently significant factor.

Concerning the affective ratings:

1. The feedback factor accounted for differences between interested-

bored ratings. Those receiving feedback rated themselves as being less bored.

2. The moderate step size variation with feedback tended to produce less boredom ratings among the middle third achievement level than the other versions but the differences were neither large nor consistent.
3. Lower third achievement level students rated the second easiest step size version rather than the easiest as most interesting.
4. Step size variation produced differences between the successful-frustrated and progress-no progress ratings both in the expected directions.

Concerning the over the summer retention-loss on the knowledge criterion scores:

1. Only prior achievement demonstrated any effect upon retention loss in that the lower third lost less. This was construed to be due to a combination of regression and "floor" effect as the lower third had little to lose.

Concerning the program error rate and time to completion:

1. All main factor effects and their interactions were significant influences upon error rate and time to complete the program.
2. The more difficult step size programs took up to less than one-half the time to complete the easier version. A saving of time always brings the question of effectiveness vs. efficiency. It may be more profitable to be concerned with time taken rather than total achieved, in that case, the larger step size versions are superior.

Concerning the common cognitive-affective criterion measures variance:

1. There was no consistent indication of any relationship existing between the cognitive measures: prior achievement test scores, error rate, knowledge test scores or application test scores, and any of the affective ratings.

Implications for New Research

The experiences with programing suggest as do Suppes and Brooks that response latency measures should be investigated as an indicator of maximum step size.

The search for intermediate level contingent generalizations for different outcomes, classes of students, and conditions in programed instruction is just beginning. Unfortunately many of those involved in the initial research efforts have turned to other problem areas leaving the major task undone. It is believed that this exodus might not have occurred if the evidence secured had fit nicely into one or another of the current learning theories.

The natural continuation of the present study is to vary independently step size and feedback with overt responding, the third basic characteristic guiding linear programing, over outcomes and classes of students. It is possible that both step size and feedback did not make more significant contributions because of the student having to make overt responses. The fact that so many students made incorrect responses seems to negate the probability. Other research also has failed to demonstrate any differences due to covert or overt responding.

Now that an indication of a step size x feedback interaction has been shown, it may be advantageous to make finer distinctions among the levels of feedback such as using certain percentage schedules. It may be that the moderate step size version is most effective with an intermittent schedule which has shown to be of value in increasing retention in schedules of reinforcement studies.

The most promising attack seems to be a systematic manipulation of the cues preceding the response, that which might be thought of as a test of the notions of cognitive theory rather than reinforcement theory. For example, a major deterrent in making a depth interpretation of the informational aspects of this study lies in the fact that three different methods were used to increase program difficulty. It is possible that removing redundant frames, review frames, and key words would have separate effects upon the multiple outcomes. The influence of these and other methods of transmitting information needs to be studied in the context of programed instruction.

Some evidence within the study points to the possibility of fitting the form of programed materials to personality traits. Not all bright students were bored by easy problems, nor did all dim students abhor a challenge. Tolerance for ambiguity levels, for example, may be quite significant in determining student output.

More concern should be paid to the concept of efficiency in learning. Many studies have overlooked, or at least played down, time as an important variable. It would have been quite interesting to have had the students repeat programs until a given time limit and then assess both learning measures. The interaction of speed and accuracy in school learning needs elaboration.

More differentiation among complex skills such as application, or transfer, in both programing and evaluation is also needed. The advent of more sophisticated statistical techniques and designs now makes it possible to probe these areas.

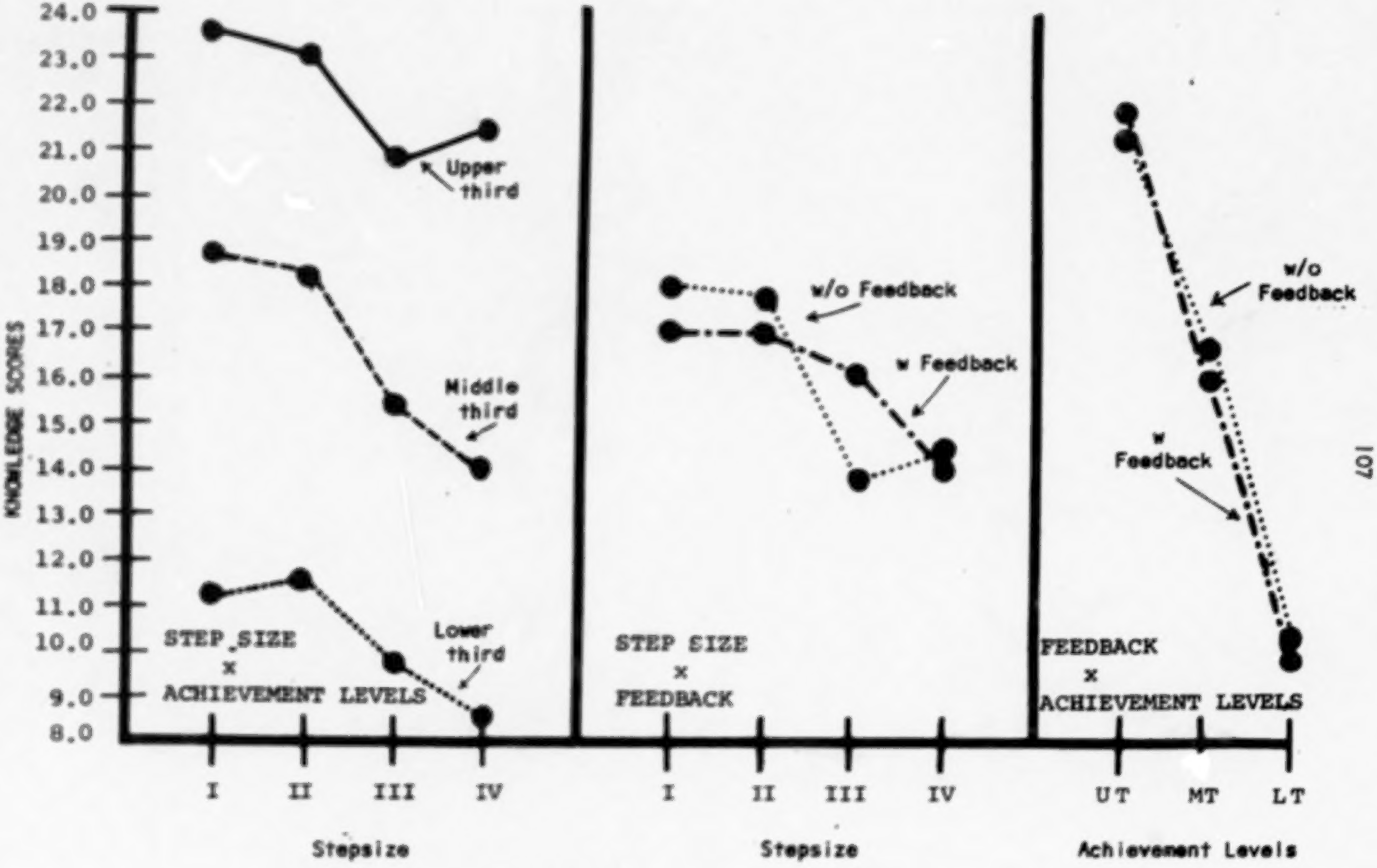
Also attempts, might be made to assess this step size x feedback interaction with:

1. different age-grade levels
2. content, both more and less abstract
3. more difficult programs
4. varying response requirements
5. shorter and longer programs
6. pictorial illustrations

Lastly, as previously mentioned, a proposal has been developed to investigate the relative effectiveness of four versions of a program which vary on the dimension, abstract to concrete, upon different achievement levels within seventh grade science students. The project is based upon the Piaget and Bruner theoretical notions that certain students need concrete representation of concepts and principles. Versions featuring laboratory experiments programed into the frames, motion pictures demonstrating these experiments, graphic illustrations representing the experiments, and a completely verbal form will be developed. The students, again divided into three achievement levels, will not only be assigned to these different versions, but also be randomly assigned to criterion instruments in the same form as the four forms of input materials.

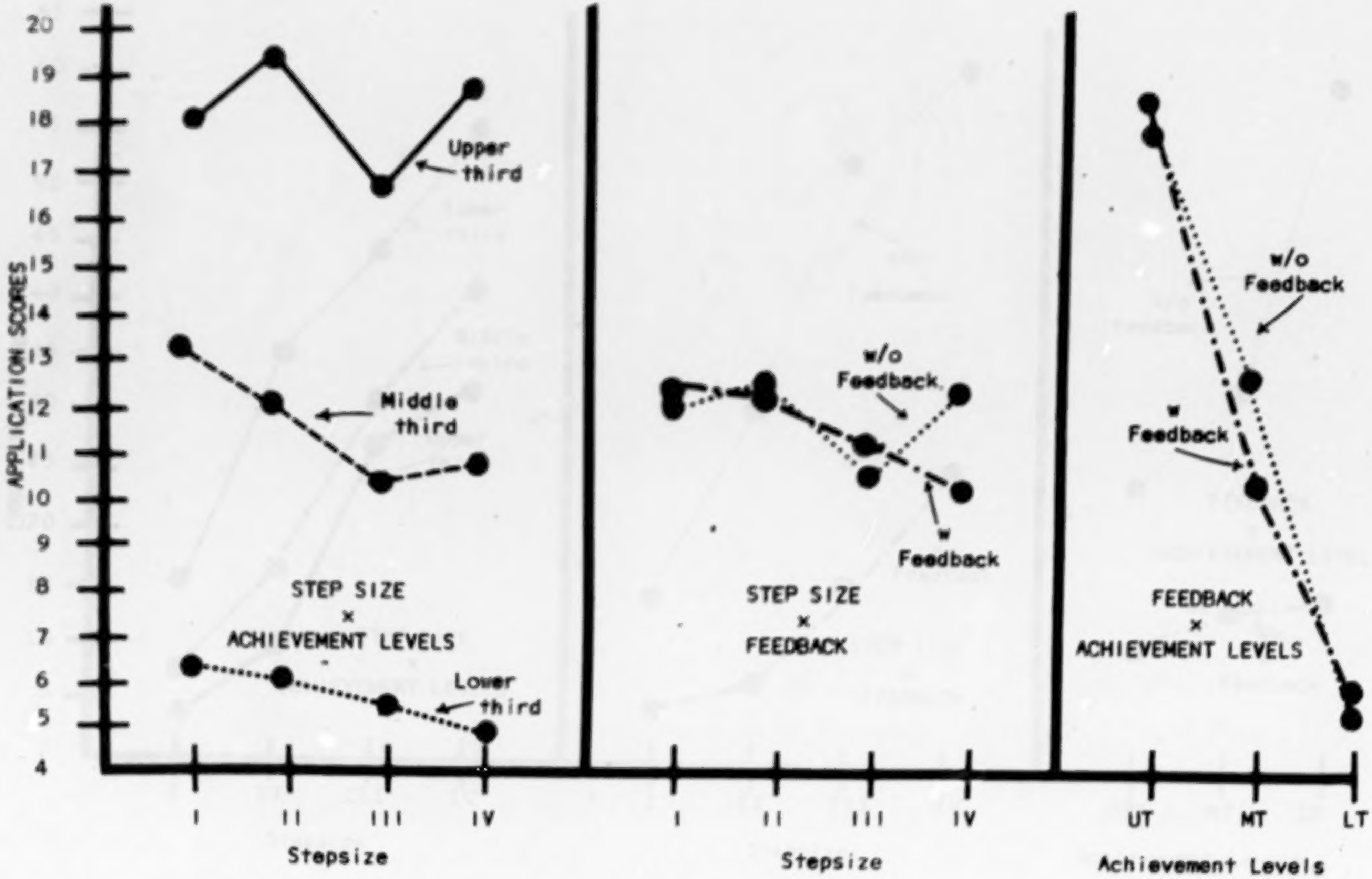
APPENDIX I

Graph	Plots of:	Page
1	The Two-Way Interactions for the Knowledge Score Means	107
2	The Two-Way Interactions for the Application Score Means	108
3	The Two-Way Interactions for the Error Rate Means	109
4	The Two-Way Interactions for the Time Means	110
5	The Two-Way Interactions for the Difficulty Levels Means	111
6	The Two-Way Interactions for the Interested-Bored Means	112
7	The Two-Way Interactions for the Alert-Careless Means	113
8	The Two-Way Interactions for the Rewarded-Punished Means	114
9	The Two-Way Interactions for the Successful-Frustrated Means	115
10	The Two-Way Interactions for the Progress-No Progress Means	116
11	The Two-Way Interactions for the Average Affect Means	117
12	The Knowledge and Application Scores for those Students Within Treatment Groups I, II, III - Variation I Showing the Triangular Distribution	118



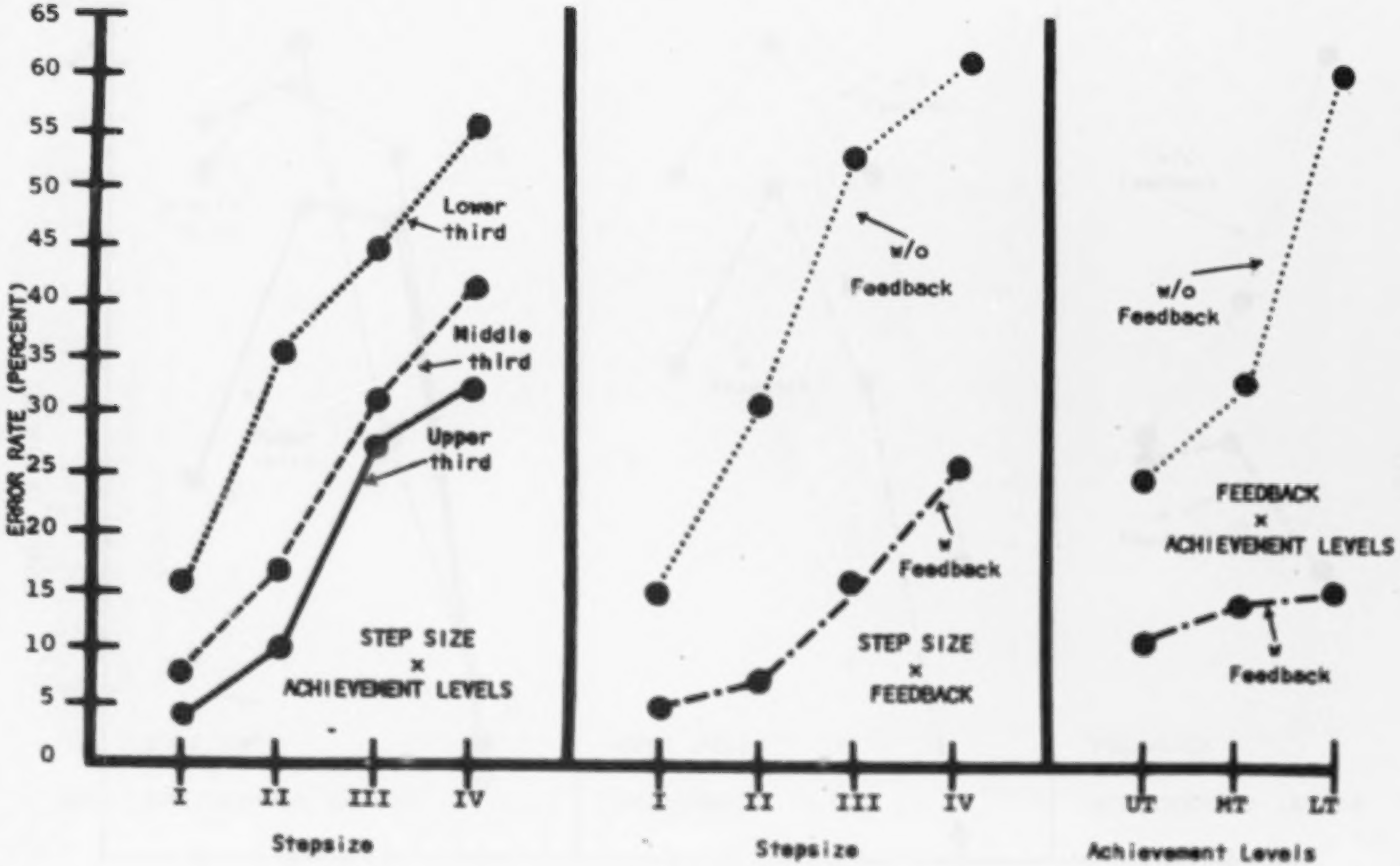
Graph #1

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE KNOWLEDGE SCORE MEANS



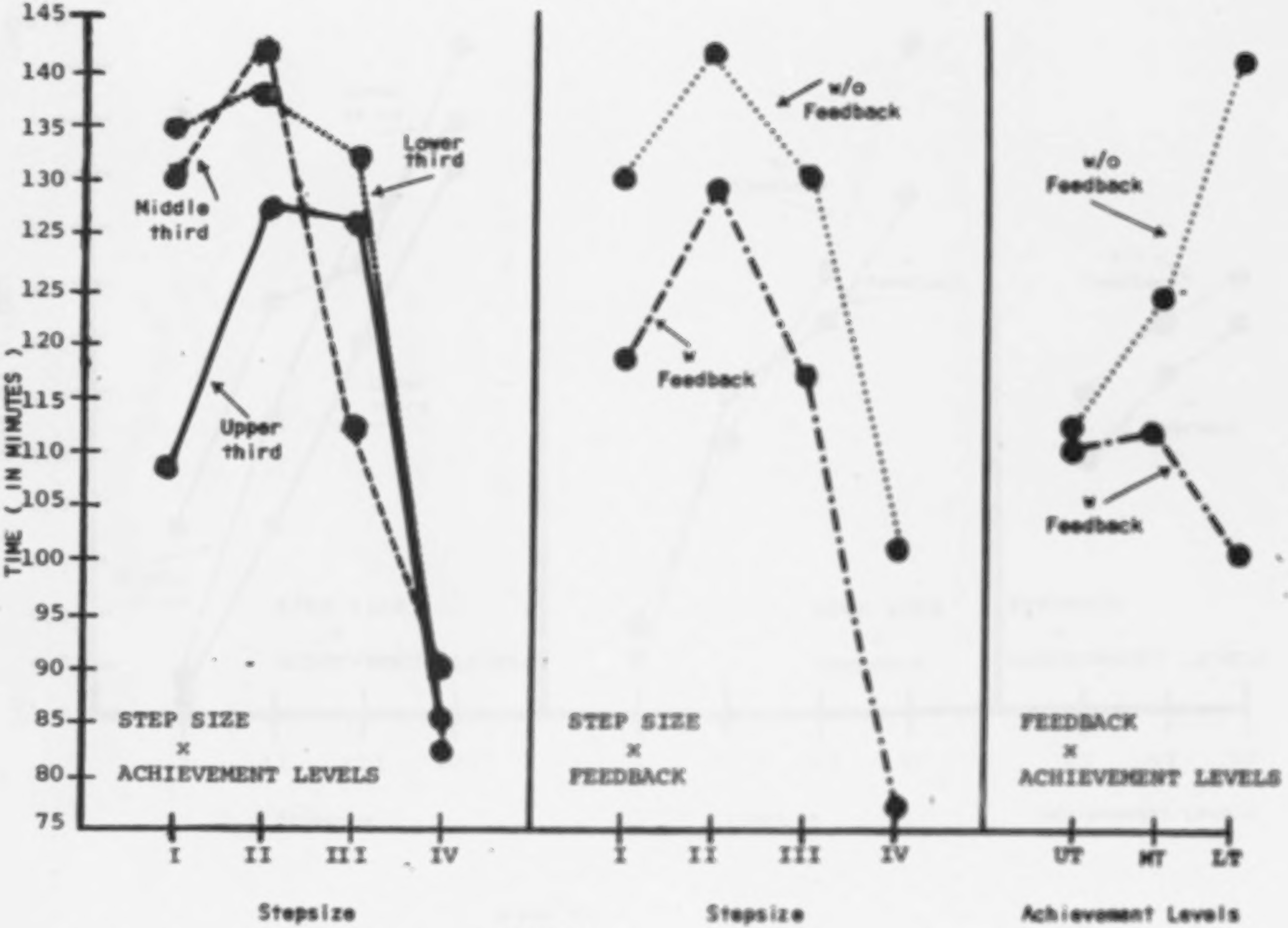
Graph #2

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE APPLICATION SCORE MEANS



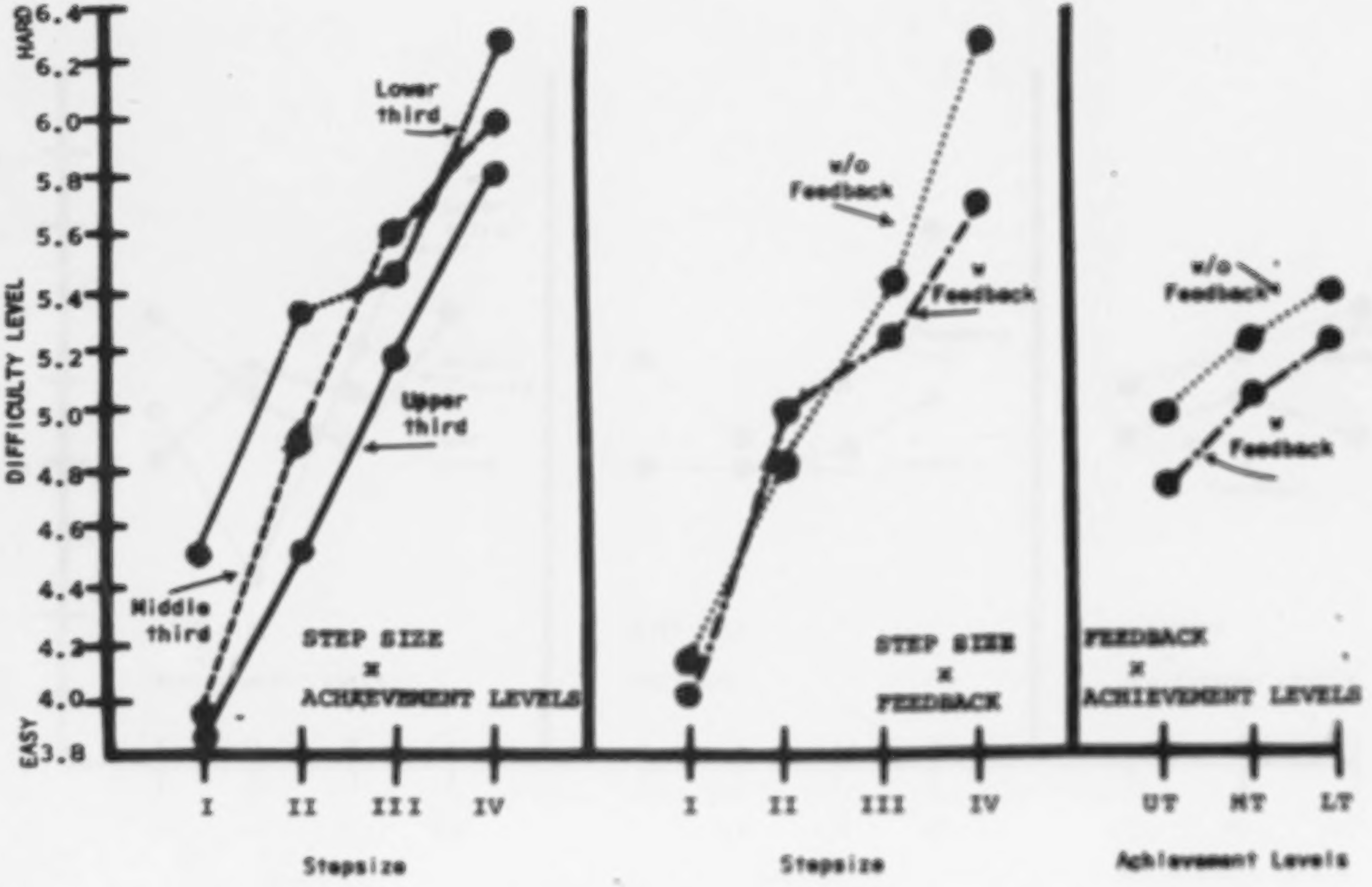
Graph #3

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE ERROR RATE MEANS



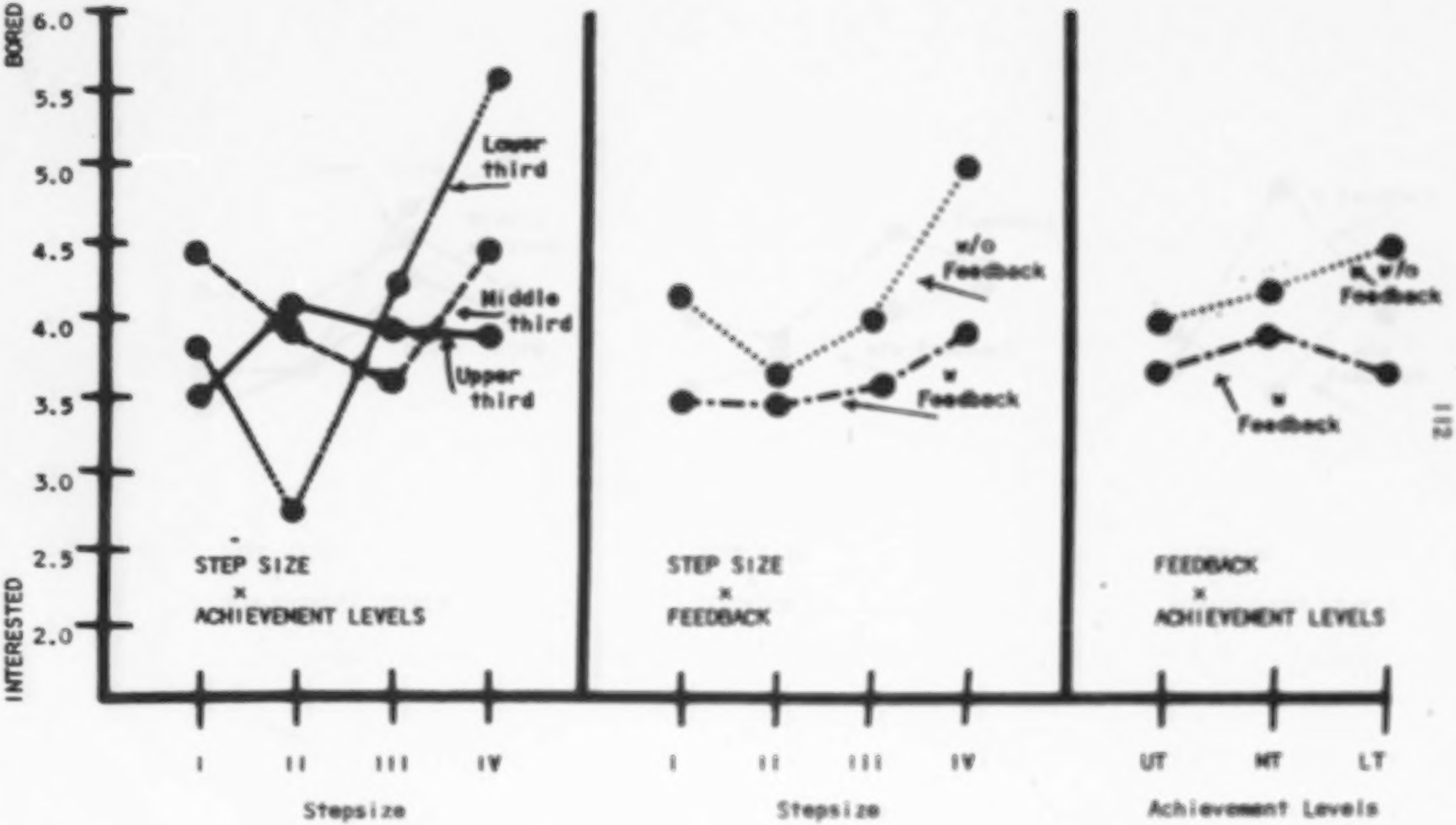
Graph #4

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE TIME MEANS



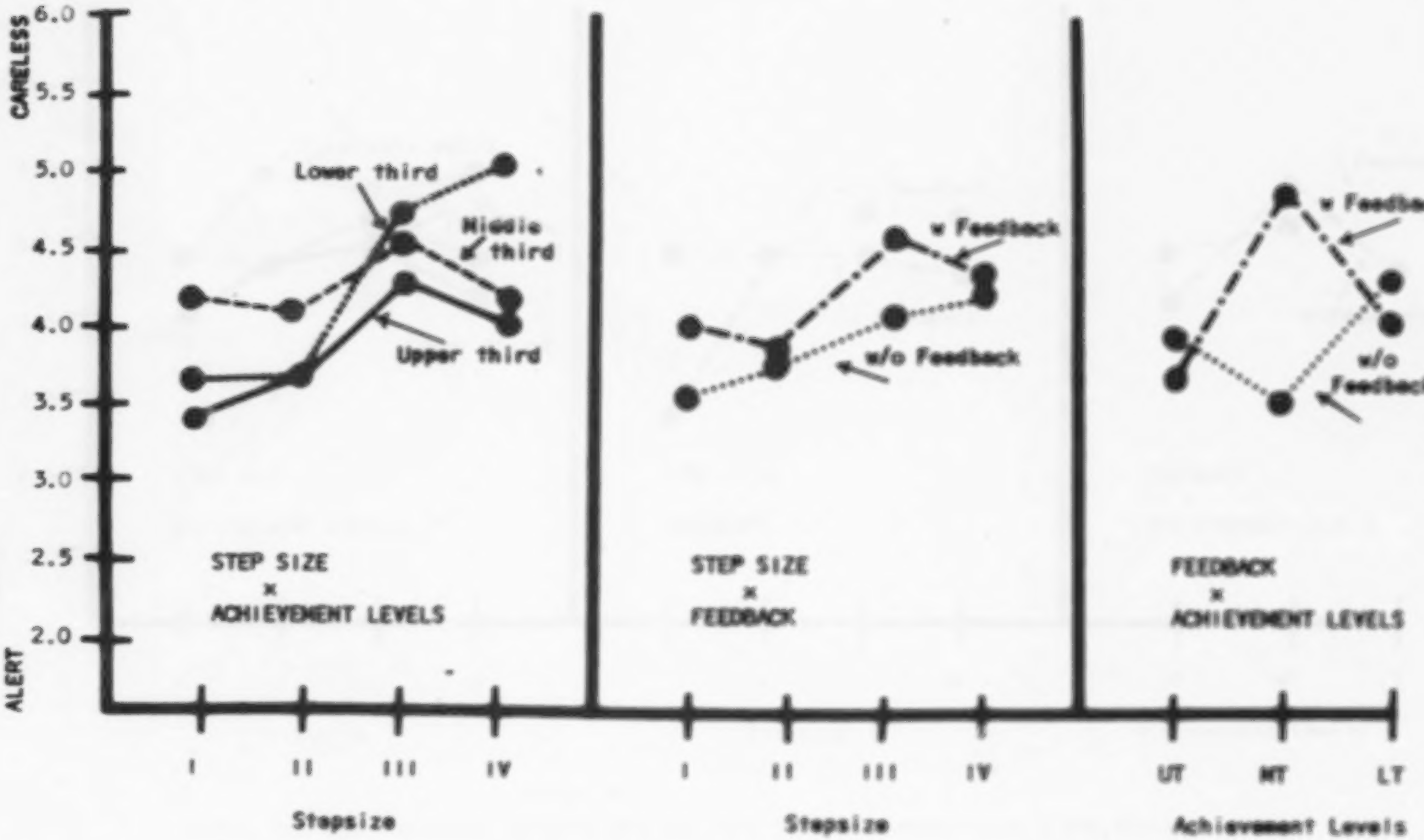
Graph #5

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE DIFFICULTY LEVELS MEANS

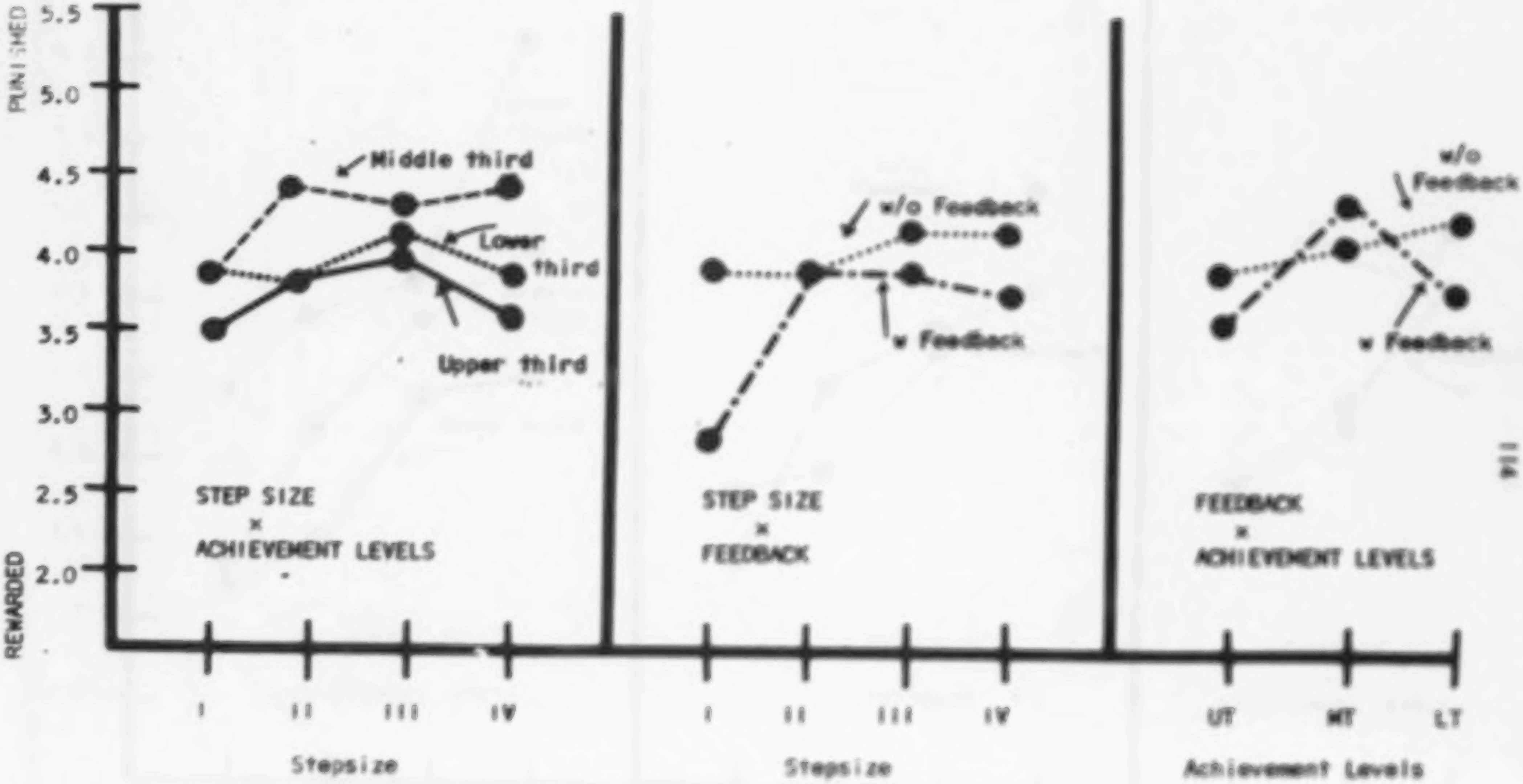


Graph #6

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE INTERESTED-BORED MEANS

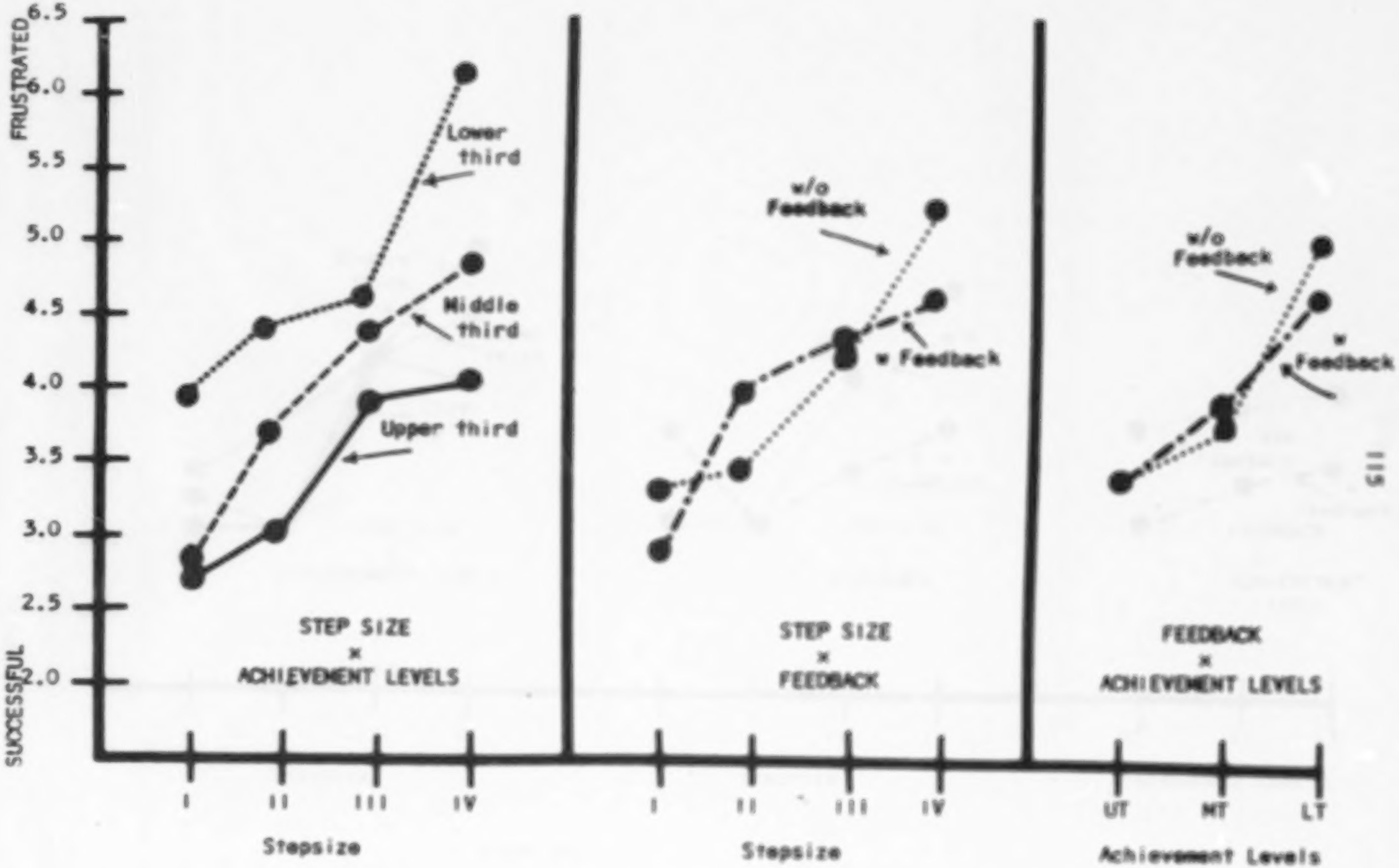


Graph #7
PLOTS OF THE TWO-WAY INTERACTIONS FOR THE ALERT-CARELESS MEANS



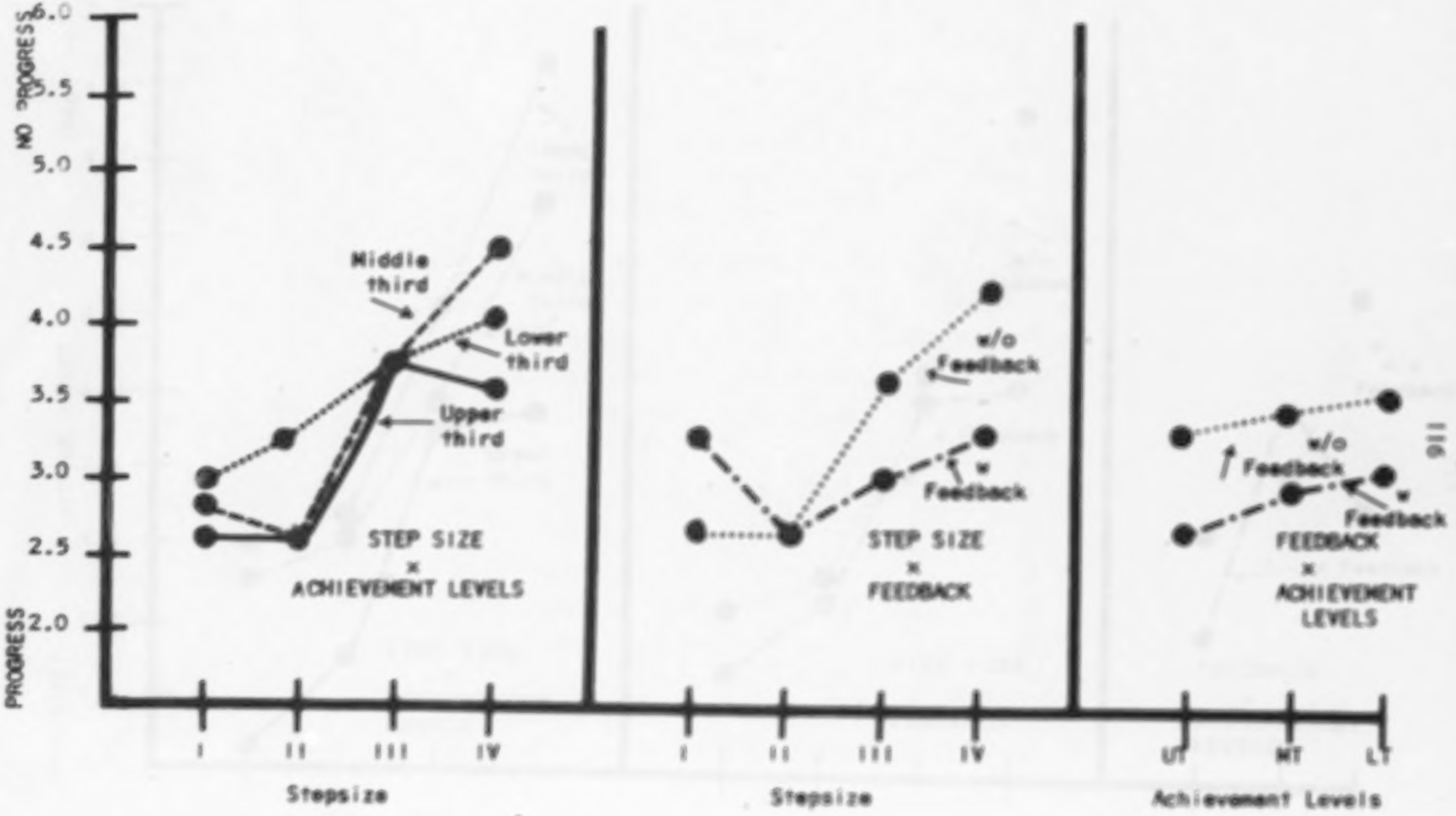
Graph #6

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE REWARDED-PUNISHED MEANS



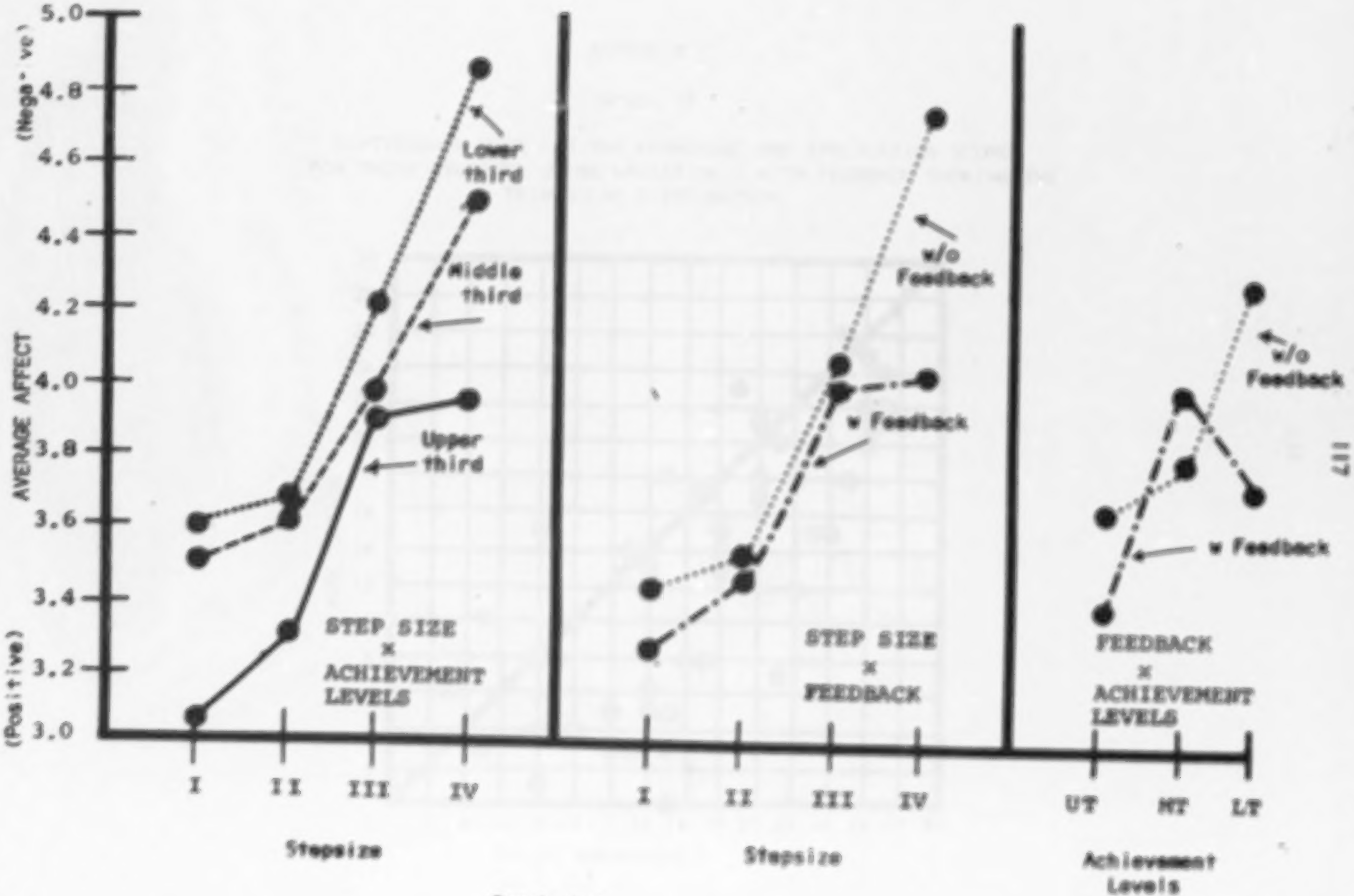
Graph #9

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE SUCCESSFUL-FRUSTRATED MEANS



Graph #10

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE PROGRESS -NO PROGRESS MEANS



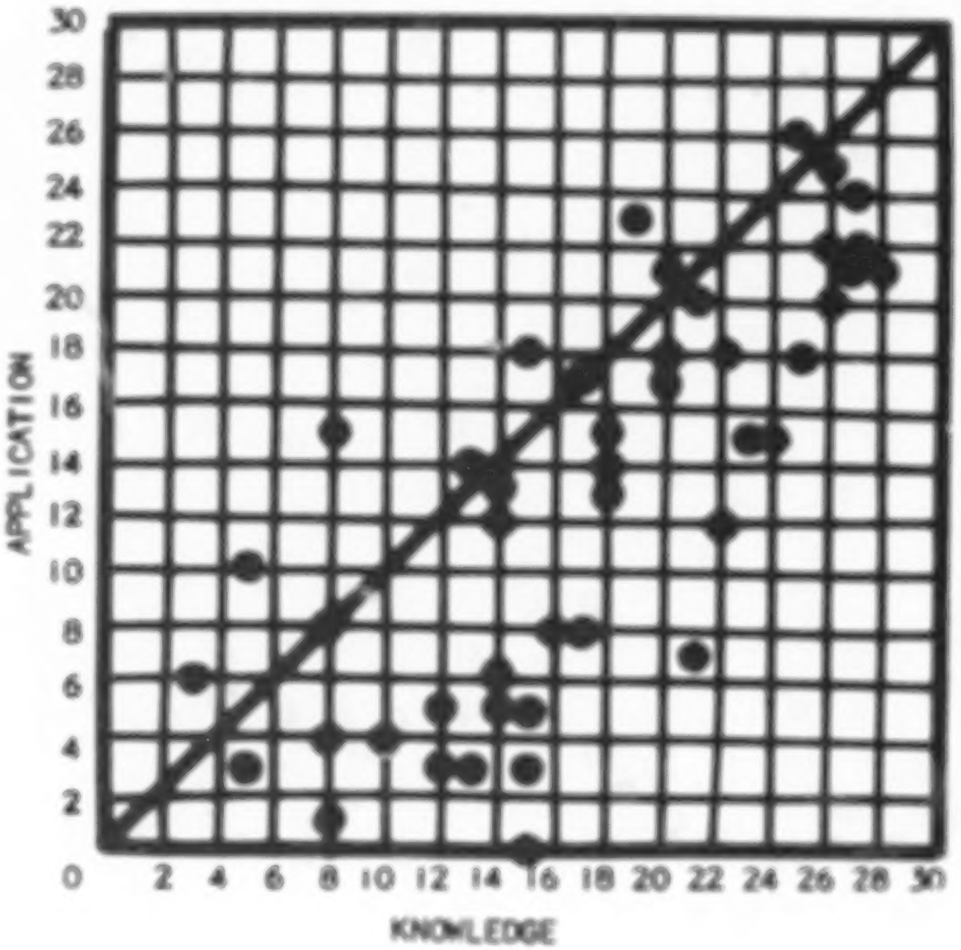
Graph #11

PLOTS OF THE TWO-WAY INTERACTIONS FOR THE AVERAGE AFFECT MEANS

APPENDIX I

Graph 12

SCATTERGRAM PLOT FOR THE KNOWLEDGE AND APPLICATION SCORES FOR THOSE STUDENTS USING VARIATION I WITH FEEDBACK SHOWING THE TRIANGULAR DISTRIBUTION



APPENDIX II

Entry Page

I Example of the Student by Frame Confidence Rating Matrix........ 120

(Students are ranked on previous science achievement)

EXAMPLE OF THE STUDENT BY FRAME CONFIDENCE RATING MATRIX
(STUDENTS ARE RANKED ON PREVIOUS SCIENCE ACHIEVEMENT)

STUDENTS	FRAMES																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	10	10	10	10	9	9	10	10	10	8	10	8	10	10	10	9	10	9	10	10	10	8
2	10	9	9	9	10	9	9	10	10	10	8	8	10	7	9	10	1	1	10	1	10	9
3	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
4	8	8	10	10	10	10	10	10	10	9	10	9	10	9	10	10	9	10	10	10	8	8
5	10	10	9	10	10	10	9	9	10	1	8	8	9	5	6	10	10	9	10	7	10	9
6	10	10	10	5	8	10	10	10	10	10	6	1	10	10	5	10	10	10	5	1	5	8
7	10	5	3	1	10	2	10	6	10	1	5	1	10	10	1	5	10	5	10	1	5	5
8	1	1	10	5	10	10	10	10	10	10	4	1	10	10	10	5	3	10	10	4	8	10
9	10	9	10	7	10	9	8	9	10	6	9	8	10	8	10	10	9	10	8	5	7	10
10	10	10	6	4	1	7	10	10	5	3	7	6	10	4	5	6	1	2	1	3	1	1
11	8	10	10	10	10	10	10	10	10	3	6	10	9	10	8	10	7	10	4	8	3	5
12	9	10	10	10	10	10	10	10	10	10	7	10	10	1	1	10	1	10	10	3	10	10
13	10	10	1	8	10	7	10	10	10	0	6	1	10	10	0	10	1	10	1	0	10	10
14	10	1	8	7	10	5	5	4	5	3	10	7	10	10	10	10	10	7	10	8	4	6
15	6	7	10	10	10	10	10	10	10	1	1	8	10	10	10	10	8	10	10	1	2	8
16	7	10	10	10	10	10	10	10	10	9	2	3	10	10	10	10	10	1	10	1	10	2
17	10	10	7	10	8	10	10	10	10	7	8	10	10	8	9	10	10	6	10	7	8	7
18	5	10	1	5	10	5	5	10	10	1	1	10	10	1	5	3	5	10	1	10	10	
19	10	10	3	10	10	10	10	10	10	5	2	10	10	10	1	10	10	7	10	10	1	8
20	10	10	10	8	8	10	10	8	10	10	7	7	10	6	8	9	9	8	7	4	6	10
21	8	1	10	9	7	10	10	10	10	10	10	6	9	5	1	4	3	7	3	5	1	1
22	0	4	0	7	5	7	6	1	8	3	10	1	9	1	7	1	4	2	2	1	7	3

APPENDIX III

Entry Page

1 Affective Questionnaire 122

2 Prior Achievement Exam 123

3 Knowledge Criterion Test 127

4 Application Criterion Test 130

QUESTIONNAIRE

Would you please answer the following questions on how you really felt about the new textbook. Be HONEST and TRUTHFUL.

(1) Were the questions in the new textbook too easy, too hard or just about the right difficulty? Circle the number which best describes how you feel. (If you don't understand, ask your teacher.)

1	2	3	4	5	6	7	8	9
much too easy	too easy	easy		just about right		hard	too hard	much too hard

(2) Circle the numbers (like you did on the first question) on each of the following scales which best describe how you REALLY felt while studying the new textbook.

1	2	3	4	5	6	7	8	9
alert								careless

1	2	3	4	5	6	7	8	9
rewarded								punished

1	2	3	4	5	6	7	8	9
making progress								making no progress

1	2	3	4	5	6	7	8	9
successful								frustrated

1	2	3	4	5	6	7	8	9
interested								bored

REMEDIAL EXAM

I. Place the next 5 words in the scientific method in order.

1. Experiment
2. Curious
3. Conclusion
4. Hypothesis
5. Observe

II. MATCHING

1. The man who said when matter is destroyed it can become energy
2. The part of an atom that spins around the center.
3. The center of an atom.
4. The part of an atom with a plus electrical charge.
5. The number of protons in one atom of an element.
6. The number of protons plus the number of neutrons.
7. The first atom smashing machine.
8. The smallest possible piece of any element.
9. Splitting, or breaking apart.
10. Combining, or coming together.

a. nucleus
b. fission
c. element
d. atomic number
e. fusion
f. neutrons
g. Einstein
h. molecule
i. cyclotron
j. Bohr
k. atomic weight
l. reactor
m. atom
n. proton
o. electron
p. Aristotle

III. Which of these can be considered "Matter". Place an "X" after the number of each substance, which is matter.

1. Chalk
2. Light
3. Oxygen
4. Coal
5. Heat
6. Iron
7. Steel
8. Nitrogen
9. Electricity
10. Sound

IV. Mark an "S" if it is a "Solution"; mark an "X" if it is a Suspension.

1. Oil and water
2. Sugar and water
3. Gas and Water
4. Starch and water
5. Salt and water

V. MATCHING

1. Centigrade
2. Motion
3. Molecular
4. Mercury
5. Alcohol
6. Kelvin
7. Temperature
8. Fahrenheit
9. Thermometer
10. Absolute

a. Pertaining to molecules
b. Used to measure heat
c. Used in most thermometers
d. British Scientist, Absolute scale
e. Household thermometer scale
f. Used in most thermometers
g. Lab thermometer scale
h. Action, movement
i. Pressure scale, positive numbers
j. Motion and rate of motion of molecules.

VI. What happens to the speed of molecules when a material is: (Use "F" if molecules move faster. Use "S" if molecules move slower.)

1. heated
2. cooled
3. contracted
4. expanded
5. changed from gas to liquid
6. changed from liquid to gas
7. changed from solid to liquid
8. changed from liquid to solid

VII. What happens to the distance between molecules when a material is: (Use "L" if distance gets larger. Use "S" if distance gets smaller.)

1. heated
2. contracted
3. expanded
4. changed from gas to liquid
5. changed from liquid to gas
6. changed from solid to liquid
7. changed from liquid to solid

[redacted]

VIII. MULTIPLE CHOICE: Mark down the letter of the Correct Answer.

1. One of these is a reptile.
   - a. eel
   - b. dinosaur
   - c. fish
   - d. frog
2. The largest fish is a:
   - a. whale
   - b. flying fish
   - c. shark
   - d. perch
3. Arthropods have:
   - a. 3 eyes
   - b. jointed appendages
   - c. fur
   - d. a diaphragm
4. Mollusk comes from a word that means:
   - a. hard
   - b. long
   - c. short
   - d. soft
5. $77^{\circ}F =$
   - a. $16^{\circ}C$
   - b. $90^{\circ}C$
   - c. $25^{\circ}C$
   - d. $27^{\circ}C$
6. A shorthand way of writing the name of a compound is called its:
   - a. equation
   - b. symbol
   - c. chemical name
   - d. formula
7. Put E if the following are elements, C if compounds, and M if mixtures.
   - a. hydrogen
   - b. water
   - c. CO_2
   - d. sugar & water

8. Purifying water by evaporating it and then condensing it again in another container, is called:
   - a. crystillation
   - b. distillation
   - c. vaporization
   - d. contamination

9. When acids and bases act upon each other they:
   - a. explode
   - b. make water
   - c. make salts
   - d. make soda

10. The following is an indication that a chemical change has occurred:
    - a. ether vaporizes
    - b. bread rises
    - c. ice cream melts
    - d. none of these

11. Carbon tetrachloride extinguishes fires by:
    - a. cooling
    - b. smothering
    - c. coating the burning material

IX. MAGIC-SQUARE EXERCISE:

Select from the answer column at the left the word which best answers or completes each of the statements at the right. Put the number of the word in the proper space in the magic-square answer box. If your answers are correct, they will form a magic square. The total of your numbers will be the same in each row across and in each column down. The four corners and the four center squares will also total this same amount.

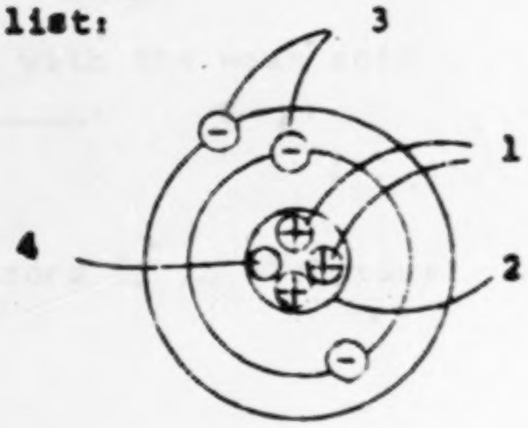
ANSWERS	STATEMENTS
1. amber	a. A north and a south pole do this
2. attract	b. A piece of iron becomes a magnet when placed in a magnetic field
3. conductor	c. Positively charged particles
4. current	d. Force that attracts iron, nickel, and cobalt
5. electrons	e. Poor transmitter of electricity
6. electroscope	f. Negatively charged particles
7. induction	g. Area of force around a magnet
8. insulator	h. Alnico is an example of this kind of magnet
9. iron	i. Charge that attracts positive electricity
10. lines of force	j. Charge that repels positive electricity
11. magnetic field	k. Used to detect an electric charge
12. magnetism	l. Good transmitter of electricity
13. negative	m. Points on a magnet where the force is concentrated.
14. neutrons	

NAME______________________________

DIRECTIONS: The following questions relate to the new textbook you have just finished studying. Read each question carefully and do your best to find the correct answer. Write your answer on the answer sheet. You are not expected to know all the answers; just do your best.

1-4. Here is a diagram of an atom.
Label the parts from the following list:

a. electron	f. cell
b. photon	g. nucleus
c. neutron	
d. molecule	
e. proton	



#1. _____ #3. _____
#2. _____ #4. _____

5. Stationary is to moving as static electricity is to __________.

6-7. The difference between dry and wet cells is in the substance in which the conductor dissolves. A wet cell has a solution and a dry cell a __________.

8. An atom that has six electrons and five protons is said to have a negative __________.

9. __________ is defined as the ability to do work.

10. __________ and electric eels are natural sources of electricity.

11. Electrical energy is important to man because he can control and harness its power to do ________.

12. An electric cell must have two __________ conductors to produce electric current.

13. The conductor that dissolves or reacts with the weak acid solution in an electric cell is __________.

14. An atom having 26 protons and 26 electrons is in electrical ______________.

15. In a dry cell, electrons flow from the zinc to the __________.

16. A charged atom must have a lack or __________ of electrons.

17. Unlike charges ___________ each other.

18. Unlike is to like as attract is to ___________.

19. The attraction between the proton and electron keep the electron in _________ around the nucleus.

20. An electrical charge that does not move is called __________.

21. A metal commonly used in a voltaic cell because it is not as reactive as zinc, is _____________.

22. When zinc atoms go into the weak acid solution, it is said that it ____________ in the solution.

23. When chemicals are mixed together and a reaction takes place, this is called ____________ __________.

24. Things that are not made by man are said to be ___________ things.

25. The plus charge (+) is called ______________.

26. The minus charge (-) is called _____________.

27. A particle that is neutral has neither a __________ or ___________ charge.

28. Anything that allows electrons to travel through it is called a _____________ of electricity.

29. The flow of electrons from atom to atom is known as an electric _____________

30. Since everything is thought to contain electrons, everything must contain ______________.

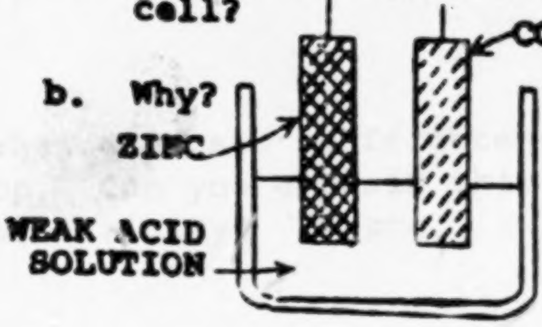
APPLICATION TEST

1. a. Tell what would happen if you were to rub a flourescent light (the long one above you in the room) with a piece of flannel or fur in a dark room.

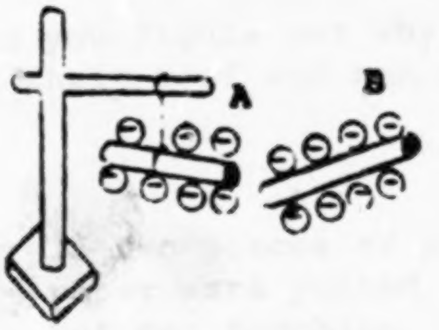
   b. Now explain why?

2. a. What's needed to complete the electron path of this voltaic cell?

   b. Why?



3.

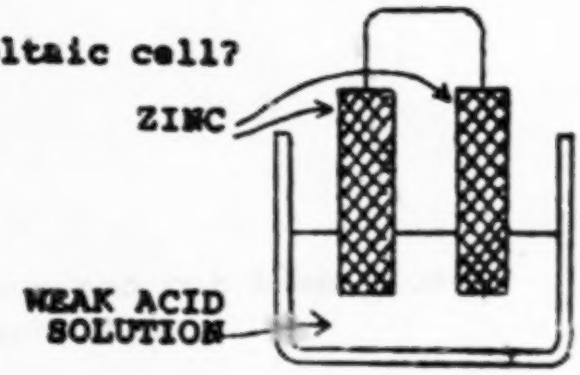


   a. What would happen to ebonite rod A when charged ebonite rod B is brought close to it?

   b. Why?

4. a. What's wrong with the following voltaic cell?

   b. How could you correct this?



5. a. If you were to rub a warm hard rubber comb through dry hair and bring it close to a thin stream of water from a faucet, what would happen?

   b. Why?

6. One of the complications with synthetic materials like nylon, orlon, and dacron, is that they tend to stick to your skin. What causes this? (2 pts)

7. a. Explain what happens when you stroke a hard rubber comb through dry hair and the hair tends to stick out in all directions.

   b. Draw a picture showing the electrical charges.

8. a. If you bring the charged comb close to the hair, what will the hair do?

   b. Why?

9. When a person is frightened, they say his hair stands straight up. Can you explain this from what you know about static electricity? (2 pts)

10. Can you figure out why people would want to put an iron pole on their roof and run this pole down into the ground? (2 pts)

11. a. If two pieces of positively charged paper were pushed close together but not touching, what would happen?

    b. Why?

12. a. If you rub two balloons on your sleeve and put them close together, they would ____________ each other.

    b. Why?

13. How would you produce electric current out of a zinc plate, copper plate, wire, and lemon. (Hint: lemon juice is a weak acid.) (2pts)

14. How would you produce electric current out of a zinc plate, copper plate, wire, and water? (2 pts)

15. You are given a glass jar, a piece of silver plate, a piece of lead plate, some weak acid solution, and a wire. You are told that the silver reacts with the acid while the lead does not. Draw a complete picture of the electric cell you can make. Show what charges would develop and the flow of electrons. (2 pts)

16. A gasoline truck drags a chain along behind it when it is moving.

a. Why do they do this?

b. Explain what happens as the truck moves along.

17. Explain where you find both static electricity and current in a thunderstorm. (2 pts)

18. If a big comb and little comb were rubbed through dry hair and then brought together:

a. which one would move? and

b. in what direction?

BIBLIOGRAPHY

Ammons, R. B. "Effects of Knowledge of Performance: A Survey and Tentative Theoretical Formulation," *Journal of General Psychology*, LIV, 1956, 279-299.

Angell, D., & Lumsdaine, A. A. *The Effects of Prompting Trials and Partial-Correction Procedures on Learning by Anticipation*. San Mateo, Calif.: American Institute for Research, September 1961 (a). Report No. AFOSR-TN-61-1343.

Ausubel, David P. *The Psychology of Meaningful Verbal Learning*. New York: Grune and Stratton, 1963.

Berlo, David K. The Process of Communication. New York: Holt, Rinehart and Winston, 1960.

Bloom, Benjamin S. (ed.). *Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain*. New York: David McKay Company, Inc., 1956.

Bloom, Benjamin S., and Krathwohl, David R., and Masia, Bertram B. *Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook II: Affective Domain*. New York: David McKay Company, Inc., 1964.

Briggs, Leslie J., and Hamilton, Nancy Russell. "Meaningful Learning and Retention: Practice and Feedback Variables," *Review of Educational Research*, XXXIV, No. 5, (December, 1964), 545-558.

Brooks, Lloyd O. "Student Response: A Guide in Learning Research." Paper presented at the American Psychological Association Convention, Los Angeles, California, September, 1964, (Mimeographed).

Bruner, Jerome S. "Some Theorems on Instruction Illustrated with Reference to Mathematics," The Sixty-third Yearbook of the National Society for the Study of Education, *Theories of Learning and Instruction*. Chicago: University of Chicago Press, 1964, 306-335.

Chansky, Norman M. "Reactions to Systems of Guiding Learning," *American Educational Research Journal*, I, No. 2, (March, 1964), 95-100.

Coulson, John E., (ed). Programmed Learning and Computer-Based Instruction. John Wiley & Sons, 1962.

Deterline, William A. "Toward an Adequate Definition of Programmed Instruction," National Society for Programed Instruction Journal, II, No. 8, 1963, 6-7.

Eigen, Lewis D. "High School Student Reactions to Programmed Instruction," Phi Delta Kappan, XLIV, No. 6, (March, 1963), 282-285.

Eigen, L. D. and Feldhusen, J. F. "Interrelationships Among Attitudes, Achievement, Reading, Intelligence, and Transfer Variables in Programmed Instruction," Educational Technology, DeCecco, John P., ed. New York: Holt, Rinehart and Winston, 1964, 376-386.

Feldhusen, J. F. and Birt, A. Journal of Educational Research, Vol. 55, 1962, 461-466.

Feldhusen, John F. "Taps for Teaching Machines," Phi Delta Kappan, XLIV, No. 6, (March, 1963), 265-267.

Ferster, Charles B. and Sapon, Stanley, M. "An Application of Recent Developments in Psychology to the Teaching of German," Teaching Machines and Programmed Learning: A Source Book, Lumsdaine, A. A. and Glaser, Robert, editors, Washington, D. C.: National Education Association, 1960, 173-185.

Fiks, Alfred I. "Some Treatment and Population Variables in Programed Instruction," Journal of Educational Psychology, Volume 55, No. 3, (June, 1964), 152-158.

Gage, N. L. "Theories of Teaching," The Sixty-third Yearbook of the National Society for the Study of Education, Theories of Learning and Instruction. Chicago: University of Chicago Press, 1964, 268-285.

Gagne, R. M. and Dick, W. "Learning Measures in a Self-Instructional Program in Solving Equations," in Educational Technology, DeCecco, John P., ed. New York, Holt, Rinehart and Winston, 1964, 315-329.

Getzels, J. W. "Creative Thinking, Problem-Solving, and Instruction," The Sixty-third Yearbook of the National Society for the Study of Education, Theories of Learning and Instruction. Chicago: University of Chicago Press, 1964, 240-267.

Goldbeck, Robert A., Briggs, Leslie Jr., Campbell, Vincent N., and Nichols, Daryl G. "Experimental Results Regarding Form of Response, Size of Step, and Individual Differences, in Automated Programs," in Coulson, John, ed. *Programmed Learning and Computer-Based Instruction*. John Wiley & Sons, 1962, 86-98.

Goldbeck, Robert A., and Campbell, Vincent H. "The Effects of Response Mode and Response Difficulty on Programed Learning." *Journal of Educational Psychology*, LIII, No. 3, 1962, 110-118.

Gotkin, Lassar G. "Cognitive Development and the Issue of Individual Differences," *Programed Instruction*, IV, No. 1, (October, 1964), 1.

Glaser, Robert. As a discussant in Hughes, J. L. (Ed.) *Programed Learning a Critical Evaluation*. Chicago: Educational Methods Incorporated, 1963.

Harris, Chester W. (ed.). *Problems in Measuring Change*. Madison, Wisconsin: The University of Wisconsin Press, 1963.

Hough, John B., and Revsin, Bernard. "Programed Instruction at the College Level: A Study of Several Factors Influencing Learning," *Phi Delta Kappan*, XLIV, (March, 1963), 286-291.

Jacobs, Paul I. "Item Difficulty and Programed Learning," *The Journal of Programed Instruction*, II, No. 2, (Summer, 1963), 21-38.

Johnson, Donald. *The Psychology of Thought and Judgment*. New York: Harper Brothers, 1955.

Klaus, David J. "Step Size and Error Rate in Programmed Instruction." Paper presented at the meetings of the American Psychological Association, September 4, 1964, (Mimeographed).

Krumboltz, John D. *Factors Affecting Difficulty Level and Criterion Performance*, (Draft of a paper presented to the American Psychological Association Annual Convention, Philadelphia, Pa., August 30, 1963).

Krumboltz, J. D. and Weisman, R. G. "The Effects of Intermittent Confirmation in Programmed Instruction," *Educational Technology*, DeCecco, John P., ed. New York: Holt, Rinehart and Winston, 1964, 291-296.

Lambert, P. and Miller, D. M., and Wiley, D. E. Journal of Educational Research, Vol. 55, 1962, 485-494.

Lipson, Shirley. "A Personal Reaction to Two Programed Textbooks," Theory Into Practice, I, No. I, (February, 1962), 7-8.

Lumsdaine, A. A. "Educational Technology, Programed Learning, and Instructional Science," The Sixty-third Yearbook of the National Society for the Study of Education, Theories of Learning. Chicago: University of Chicago Press, 1964, 371-401.

Lumsdaine, A. A. (Ed.) Student Response in Programmed Instruction. Washington, D.C.: National Academy of Sciences -- National Research Council, 1961, Publication 943.

McClelland, David C., Atkinson, J. W., Clark, R. A., and Lowell, E. L. The Achievement Motive. New York: Century-Crofts Inc., 1953.

Michael, D. N. and Maccoby, N. "Factors Influencing Verbal Learning from Films under Varying Conditions of Audience Participation," Educational Technology, DeCecco, John P., ed. New York: Holt, Rinehart and Winston, 1964, 267-275.

Moore, J. William, and Smith, Wendell I. "Knowledge of Results in Self-Teaching Spelling," Psychological Reports, IX, 1961, 717-726.

Naumann, T. A. Journal of Programmed Instruction, Vol. I, 1962, 9-18.

Porter, D. "Some Effects of Year Long Teaching Machine Instruction," Automatic Teaching: The State of the Art, Galanter, E. H., ed. New York: John Wiley and Sons, 1959, p. 85-90.

Read, Edwin A. "Problems in the Use of Programed Instruction in the Continuous Progress School," Programed Instruction, IV, No. I, (October, 1964), 9.

Reed, J. E. and Hayman, J. W. Journal of Educational Research, Vol. 55, 1962.

Roth, Robert Howard. "Student Reactions to Programed Learning," Phi Delta Kappan, XLIV, No. 6, (March, 1963), 278-281.

Scheffe, Henry. The Analysis of Variance. New York: John Wiley and Sons, Inc., 1959.

Shay, Carleton B. "Relationship of Intelligence to Size of Step on a Teaching Machine Program" *Journal of Educational Psychology*, LII, No. 2, 1961, 91-103.

Skinner, B. F. "The Science of Learning and the Art of Teaching," *Teaching Machines and Programmed Learning: A Source Book*, Lumsdaine, A. A. and Glaser, Robert, editors, Washington, D.C.: National Education Association, 1960, 99-113.

Smith, M. Daniel. "Let's Not Talk About 'Step Size'," Earlham College, (Mimeographed).

Smith Wendell I., and Moore, J. William. *Programed Learning: Theory and Research*, New York: D. VanNostrand Company, Inc., 1962.

Strong, Paschal N., Jr. "Research Accomplishments and Needs in Programmed Instruction," *Trends In Programmed Instruction*, Ofresh, Gabriel D., ed. Washington, D.C.: National Education Association, 1964, 224-230.

Suppes, Patrick. "Modern Learning Theory and the Elementary-School Curriculum," *American Educational Research Journal*, I, No. 2, (March, 1964), 79-93.

Thorndike, Edward L. *Educational Psychology*, *Vol. II*, *The Psychology of Learning*, Teachers College, Columbia University. New York, 1913.

Travers, R. M. W., and Van Wagenen, R. Keith, and Haygood, Danielle H., and McCormick, Mary. "Learning as a Consequence of the Learner's Task Involvement under Different Conditions of Feedback," *Journal of Educational Psychology*, Volume 55, No. 3, (June, 1964), 167-173.

Wallace, John. "Concept Dominance, Type of Feedback, and Intensity of Feedback as Related to Concept Attainment," *Journal of Educational Psychology*, Volume 55, No. 3, (June, 1964), 159-166.

Wendt, Paul R. and Rust, Grosvenor C. "Programed Instruction for Transfer to the Real Life Situation," *Phi Delta Kappan*, XLIV, No. 6, (March, 1963), 273-277.

Winer, B. J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill Book Co., Inc., 1962.

Wittrock, M. C. and Twelker, P. A. "Prompting and Feedback in the Learning Retention, and Transfer of Concepts," *British Journal of Educational Psychology*, Fall 1964, 10-18.